

Audio-Gamification:

How Gamification Can Improve the Use of
Intelligent Virtual Assistants

Paula Bräuer



C | A | U

Christian-Albrechts-Universität zu Kiel

Technische Fakultät

Audio-Gamification: How Gamification Can Improve the Use of Intelligent Virtual Assistants

Paula Bräuer

Dissertation

zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)

der Technischen Fakultät
der Christian-Albrechts-Universität zu Kiel
eingereicht im Jahr 2024

Erste Gutachterin: Prof. Dr. Isabella Peters

Zweiter Gutachter: Prof. Dr. Thomas Voit

Datum der mündlichen Prüfung 22.03.2024

Abstract

Gamification is an established approach in the field of human-computer interaction that aims to motivate users by using game design elements in a non-game context. This concept has proven successful in many areas, including the penetration of technologies such as fitness trackers into the market. In recent years, intelligent virtual assistants (IVAs) like Amazon Alexa or Google Assistant have become increasingly popular, offering a new type of human-machine interaction. The exchange of information with an IVA through voice offers several advantages. Firstly, it requires no physical touch, eliminating the need for a keyboard or touch device. Secondly, information can be accessed without the need to look at a screen, which is particularly useful when multitasking. However, studies have shown that the use of IVAs decreases over time. It appears that interaction with an IVA is not yet attractive enough, and there is a lack of incentives to make the interaction motivating for the user. Previous work on the motivating effect of gamification in various contexts shows that visual game design elements are commonly used for implementation. However, many smart speakers, which are among the most commonly used end devices for IVAs, do not have a screen. This raises the research question of whether and how gamification can be used in the context of IVAs without visual support. This thesis addresses this research gap by conducting four studies to investigate different aspects of this question. The first study aimed to determine if individual game design elements can have a motivating effect. The second study, a laboratory experiment, investigated the impact of gamification without visual support on motivation and performance. Study three built upon this by examining the effects of different audio gamification concepts on motivation and performance in a field experiment. Finally, Study Four conducts an analysis of IVA games, extracting game design elements to gain insight into which elements might be suitable for implementing gamification for IVAs. In summary, this dissertation demonstrates the circumstances under which it is possible to implement gamification without visual support in the context of IVAs.

Zusammenfassung

Ein Ansatz, der sich im Bereich der Human Computer Interaction etabliert hat, um Nutzer zu motivieren, ist Gamification. Unter Gamification versteht man die Verwendung von Spiel Design Elementen in einem spielfremden Kontext. Das Konzept hat sich in vielen Bereichen bewährt und Technologien wie Fitnesstrackern zur Marktdurchdringung verholfen. Intelligente virtuelle Assistenten (IVA) wie Amazon Alexa oder Google Assistant bieten eine neue Art der Mensch-Maschine-Interaktion, die sich in den letzten Jahren immer stärker durchgesetzt hat. Die Möglichkeit, Informationen mit einem IVA per Sprache auszutauschen, bietet verschiedene Vorteile. Zum einen ist für die Eingabe keine Berührung notwendig, es muss also weder eine Tastatur noch ein Touchgerät verwendet werden. Zum anderen können Informationen abgerufen werden, ohne dass der Blick auf einen Bildschirm gerichtet werden muss, was insbesondere beim Multitasking von Vorteil ist. Dennoch zeigen verschiedene Studien, dass die Nutzung von IVAs mit der Zeit abnimmt. Es scheint also, dass die Interaktion mit einem IVA derzeit noch nicht attraktiv genug ist und es an Anreizen fehlt, die die Interaktion für den Nutzer motivierend genug machen. Bei der Betrachtung bisheriger Arbeiten zur motivierenden Wirkung von Gamification in verschiedenen Kontexten fällt auf, dass zur Umsetzung meist mit visuellen Spiel Design Elementen verwendet werden. Viele Smart Speaker, die zu den meistgenutzten Endgeräten für IVA zählen, verfügen jedoch nicht über einen Bildschirm. Aus diesen Erkenntnissen leitet sich die Forschungsfrage ab, ob und wie Gamification auch im Kontext von IVAs ohne visuelle Unterstützung eingesetzt werden kann. Diese Forschungslücke wird im Rahmen dieser Arbeit adressiert. Dazu wurden vier Studien durchgeführt, die unterschiedliche Teilaspekte dieser Fragestellung untersuchen. In der ersten Studie wurde untersucht, ob eine motivierende Wirkung bereits durch den Einsatz einzelner Spiel Design Elemente erzielt werden kann. Die zweite Studie ist ein Laborexperiment, in dem erstmals untersucht wird, wie sich Gamification ohne visuelle Unterstützung auf Motivation und Leistung auswirkt. Darauf aufbauend wird in Studie drei in einem Feldexperiment untersucht, wie sich verschiedene Audio-Gamification-Konzepte auf Motivation und Leistung auswirken. Abschließend wird in Studie vier eine Analyse von IVA-Spielen durchgeführt, bei der Spiel Design Elemente aus den Spielen extrahiert werden, um einen Einblick zu erhalten, welche Elemente für die Implementierung von Gamification für IVAs geeignet sein könnten. Zusammenfassend zeigt diese Dissertation, unter welchen Umständen es möglich ist, Gamification auch ohne visuelle Unterstützung im Kontext von IVAs umzusetzen.

List of Abbreviations

| | |
|--------------|------------------------------------|
| AI | Artificial Intelligence |
| AWS | Amazon Web Services |
| CUI | Conversational User Interface |
| GUI | Grafical User Interface |
| HCI | Human-Computer Interaction |
| HTTPS | Hypertext Transfer Protocol Secure |
| IVA | Intelligent Virtual Assistant |
| L1 | native speaker |
| L2 | second language speaker |
| ML | Machine Learning |
| RQ | Research Question |
| SDT | Self Determination Theory |
| SSML | Speech Synthesis Markup Language |
| VUI | Voice User Interface |

List of Figures

| | |
|--|----|
| Figure 1.1 Process of developing the research idea and questions | 4 |
| Figure 1.2 Relationship of the four research questions to address the research gap..... | 5 |
| Figure 1.3 Structure of this Thesis..... | 8 |
| Figure 1.4 Model of game design elements by Werbach and Hunter adapted from (Werbach & Hunter, 2012)..... | 14 |
| Figure 1.5 The Periodic Table of Gamification Elements by (Marczewski, 2018)..... | 15 |
| Figure 1.6 IVAs are used especially on so-called smart speakers. Shown are an Amazon Echo Dot 4th generation (left), 3rd generation (center), and an Amazon Echo Show (right)..... | 17 |
| Figure 1.7 Generalized Dialogue Architecture (following M. McTear et al., 2016)..... | 19 |
| Figure 1.8 Illustration of the concept of sessions based on the JOVO documentation | 26 |
| Figure 2.1 (Study1) Sample question from the narrative condition..... | 39 |
| Figure 2.2 (Study1) Sample question from the feedback condition..... | 39 |
| Figure 2.3 (Study 1) Sample question of the progress bar condition..... | 41 |
| Figure 2.4 (Study 1) Frequency distribution of answered questions per condition/group for the individual game design elements..... | 42 |
| Figure 2.5 (Study 1) Frequency distribution of answered questions per condition/group for the combined game design elements..... | 43 |
| Figure 3.1 (Study 2) Example of a dialogue of a subject in the control group..... | 57 |
| Figure 3.2 (Study 2) Example of a dialogue between a subject and a gamified version of the Alexa skills..... | 57 |
| Figure 3.3 (Study 2) Description of the cooperative and competitive ranking in the three gradations negative, neutral, and positive..... | 58 |
| Figure 4.1 (Study 3) Example dialog in the control group when starting the Alexa skill for the first time..... | 76 |
| Figure 4.2 (Study 3) The flowchart illustrates the interaction with the Alexa skill in the control group..... | 76 |
| Figure 4.3 (Study 3) General flowchart illustrates the interaction with the Alexa skill in the experimental groups (achievements and leaderboard)..... | 77 |
| Figure 4.4 (Study 3) Example of a dialog in which the user calls up the leaderboard..... | 78 |
| Figure 4.5 (Study 3) Example dialog for a user who receives an achievement and starts a challenge..... | 78 |

Figure 5.1 (Study 4) Flowchart of the Alexa skill selection.....92

Figure 9.1 The pieces of the puzzle that fit together to illustrate the connection between the four research questions.....111

List of Tables

| | |
|--|----|
| Table 1.1 (Study 1) List of all badges with the condition for unlocking them and number of subjects who received them..... | 40 |
| Table 1.2 (Study 1) Intervals and questions necessary to answer for progress in the progress bar condition..... | 41 |
| Table 1.3 (Study 1) Number, mean value, standard deviation (SD), p-value, and the number of subjects with all questions answered per condition..... | 43 |
| Table 2.1 (Study 2) Overview of the degree of previous experience with IVAs of the subjects..... | 59 |
| Table 2.2 (Study 2) Number of subjects, as well as the mean, median, and standard deviation of the adjusted required time per experimental group..... | 59 |
| Table 2.3 (Study 2) Mean and standard deviation for all three motivational items (General motivation/Motivation to work faster/Feeling of fun) per experimental group..... | 60 |
| Table 2.4 (Study 2) Results of the Mann-Whitney-U-Test and effect size for the comparison between the control group and the competitive group regarding the three items for measuring motivation..... | 60 |
| Table 2.5 (Study 2) Results of the Mann-Whitney-U-Test and effect size for the comparison between the control group and the cooperative group concerning the three items for measuring motivation..... | 60 |
| Table 2.6 (Study 2) Results of the Mann-Whitney-U-Test for the comparison between the cooperative and the competitive group regarding the three items for measuring motivation..... | 60 |
| Table 2.7 (Study 2) Correlation matrix of the additional statistical analyses. P-values are two-tailed..... | 60 |
| Table 2.8 (Study 2) Number of subjects, as well as the mean, median, and standard deviation of the adjusted required time per experimental group and gender..... | 60 |
| Table 3.1 (Study 3) List of all achievements with their title and description..... | 79 |
| Table 3.2 (Study 3) Mean, standard derivation, and median of the sum of processed vocabulary, correct answers, and time in seconds per user..... | 80 |
| Table 3.3 (Study 3) Mean, standard derivation, and median ratio of correct vocabulary to the sum of answered vocabulary..... | 81 |
| Table 4.1 (Study 4) List of elements according to Werbach and Hunter and the frequency at which they were found in the 29 Alexa games analyzed..... | 93 |

Acknowledgments

Zunächst möchte ich mich bei meiner Betreuerin Isabella Peters bedanken. Sie hat mir sowohl bei der gemeinsamen Projektarbeit als auch bei der Betreuung dieser Dissertation viel Freiheit gegeben, so dass ich meine eigene Forschung und Ideen verfolgen konnte und bei Bedarf jederzeit auf ihr Feedback setzen konnte. Besonders beim Zusammenfügen der einzelnen Teile in der Schlussphase war sie eine große Hilfe. Ebenso möchte ich mich bei Thomas Voit bedanken, der mir wichtige Denkanstöße gegeben hat und mich zum richtigen Zeitpunkt zur Fertigstellung der Arbeit ermutigt hat.

Athanasios Mazarakis nimmt einen besonders wichtigen Platz in meinem wissenschaftlichen Werdegang ein. Er hat mir den Weg in die Wissenschaft gezeigt, mich für die Forschung begeistert und mir die Chance gegeben, in diesem Bereich zu arbeiten und meine eigenen Ideen zu verfolgen. Ohne ihn würde es diese Arbeit heute nicht geben. Sein Beitrag als Co-Autor und seine Beteiligung an der Entwicklung und Diskussion neuer Ideen waren für mich von unschätzbarem Wert.

Auch meinen Kollegen Isabelle Dorsch, Maria Henkel, Armin Jacob, Steffen Lemke und Lennart Perrey möchte ich an dieser Stelle für die angenehme Arbeitsatmosphäre, die schönen gemeinsamen Mittagessen und Kaffeepausen danken. Der Austausch mit euch hat die Arbeit sehr beschleunigt und die Arbeitstage im Büro sehr bereichert. Ebenso danke ich allen anderen aktuellen und früheren Kolleg:innen an der ZBW, die im Laufe meiner Promotion für mich da waren, ob zu Beginn (besonderer Dank an Falk Bösch) oder in der Endphase (besonderer Dank an Alisa Held und Robert Bock).

Ich danke außerdem meinen Freundinnen Milena Niemeyer, Leonie Schmiedel, Maggie Berg und Fritzi Menke, die immer für mich da waren und mir geholfen haben, diesen Weg bis zum Ende zu gehen.

Einen großen Dank möchte ich meinen Eltern, meiner Schwester und Bär aussprechen, die mich immer unterstützt, zum Weitermachen motiviert und sich mit mir über jeden Erfolg gefreut haben. Auch in Momenten des Zweifels wart ihr für mich da und habt mich aufgebaut - dafür bin ich zutiefst dankbar. ☺

Table of Contents

| | |
|---|------|
| Abstract | ii |
| Zusammenfassung..... | iii |
| List of Abbreviations..... | iv |
| List of Figures | v |
| List of Tables..... | vii |
| Acknowledgments..... | viii |
| 1. Introduction and Motivation | 1 |
| 1.1. Publication Record | 3 |
| 1.2. Research Questions | 3 |
| 1.3. Outline | 8 |
| 2. Related Work | 10 |
| 2.1. Gamification..... | 10 |
| 2.1.1. Motivation Theories | 12 |
| 2.1.2. Game Design Elements | 13 |
| 2.2. Intelligent Virtual Assistants | 16 |
| 2.2.1. Benefits of IVAs | 17 |
| 2.2.2. Technical Implementation of IVAs..... | 18 |
| 2.2.3. Introduction to Alexa Skill Development | 21 |
| 2.2.4. Challenges in The Design of IVA Applications | 27 |
| 2.3. (Video-)gaming Without Graphics..... | 29 |
| 3. Impact of Individual Game Design Elements | 32 |
| 3.1. Preface to Study 1 | 32 |
| 3.2. Study 1 (publication) | 33 |
| 4. Audio-Gamification Lab Experiment | 50 |
| 4.1. Preface to Study 2..... | 50 |
| 4.2. Technical Aspects of Study 2 | 51 |
| 4.3. Study 2 (publication) | 51 |

| | |
|--|-----|
| 5. Audio-Gamification Field Experiment | 66 |
| 5.1. Preface to Study 3..... | 66 |
| 5.2. Technical Aspects of Study 3 | 67 |
| 5.3. Study 3 (publication)..... | 69 |
| 6. Game Design Elements from Alexa Games | 88 |
| 6.1. Preface to Study 4..... | 88 |
| 6.2. Study 4 (publication)..... | 89 |
| 7. General Discussion | 100 |
| 7.1. Theoretical Implications..... | 100 |
| 7.2. Practical Implications | 104 |
| 7.3. Future Research Directions | 107 |
| 8. Conclusion | 110 |
| References..... | 113 |
| Appendix..... | 130 |

CHAPTER ONE

Introduction and Motivation

Controlling computers by voice has long been a valued method in many science fiction movies, such as the “Star Trek” television series or the movie “Her.” At any time, whether in an elevator on a spaceship or at home on the sofa, the movie characters can use speech to request information, ask the computer for assistance, or just have a nice chat with the virtual voice. On the starship Enterprise, for example, you can simply say “Tea, Earl grey, hot” to order a hot black tea from the replicator without having to click through any selection menus. With the introduction of Apple’s Siri in 2011, voice interaction was made possible for the first time to a large number of users in the real world via their smartphones (Guzman, 2019). Talking to Siri on the iPhone was a new experience that impressed many users despite the relatively small number of possible requests. A little later, Amazon Alexa, the next intelligent virtual assistant, was introduced in 2014 (Bouزيد & Ma, 2022). Unlike Siri on the smartphone, it was possible to communicate with Alexa via a smart speaker that is activated without having to press a button first. Furthermore, the user no longer needed to stand directly next to the device but could communicate with Alexa over a short distance across a room. Whether on a smartphone or at home on a smart speaker, the prevalence of Intelligent Virtual Assistants (IVAs) in our daily lives is constantly increasing. The global market for smart speakers has grown steadily in recent years to reach \$16.13 billion by 2022, and forecasts are for further growth (Laricchia, 2023). Despite the now widespread adoption of voice assistants, use is mostly limited to short requests with little interaction, such as playing music or asking about the weather (McTear, 2021). There seems to be a lack of incentives for extensive interaction with the voice assistant. As studies have shown, both usage (Lopatovska et al., 2019) and satisfaction (Luger & Sellen, 2016) decrease over time as users become aware of the

limitations of IVAs, such as requests that are not properly understood, or requests that are considered simple but for which functions have not yet been implemented.

One way to motivate users is through gamification, often defined as the use of game design elements in non-game contexts (Deterding et al., 2011). Currently, gamification is commonly utilized to encourage consistent or extended usage of software applications. Smartphone apps from a wide variety of fields such as wellbeing (e.g., Headspace), education (e.g., Duolingo), and transportation (e.g., Uber) show the user's progress or offer the possibility to compare themselves with other users (Fish & Saul, 2019; Shortt et al., 2023; Vasudevan & Chan, 2022). For this purpose, various game design elements are applied, such as leaderboards, levels or badges, digital artifacts that are awarded to the user for completing certain activities (Antin & Churchill, 2011). However, all of these software applications work with visual game design elements to implement gamification. What distinguishes interaction via voice is that the visual representation of information is not mandatory. By communicating via spoken words, information can be conveyed without interrupting the current activity, for example, by looking at a screen (Bouzid & Ma, 2022). At first glance, the idea of implementing a concept like gamification, which traditionally relies heavily on visual components, within the context of IVAs might appear counterintuitive. Gamification is often associated with interactive visual interfaces that engage users through graphical elements like progress bars, and badges. The idea of exploring whether and how gamification could be implemented for IVAs was broadly outlined (Rackwitz & Niesenhaus, 2018). A number of ideas were put forward, from narratives in which an IVA acts as the narrator, to concepts for social mechanisms such as competition with other users. Overall, the idea was deemed exciting but challenging to implement. However, it's worth noting that the concept of gaming is not inherently tied to screens or visual cues. Board games, for example, have effectively engaged players for generations without the use of screens. This prompts the question: if gaming experiences can be translated into non-visual formats like board games, could the principles of gamification similarly be adapted to voice-based interactions with IVAs? This proposition challenges the assumption that gamification must rely solely on visual elements, and suggests that audio cues and interaction could be utilized to create engaging and motivating experiences in this specific context. Gamification for IVAs thus represents a pioneering research gap, the investigation of which may well have significant potential for research and practice.

1.1. Publication Record

This thesis is based on four publications, which have been published in peer-reviewed scientific journals and presented at a scientific conference:

Mazarakis, A., & Bräuer, P. (2023). Gamification is Working, but Which One Exactly? Results from an Experiment with Four Game Design Elements. *International Journal of Human-Computer Interaction*, 39(3), 612–627. <https://doi.org/10.1080/10447318.2022.2041909>

Bräuer, P., & Mazarakis, A. (2022). “Alexa, Can We Design Gamification Without a Screen?”—Implementing Cooperative and Competitive Audio-Gamification for Intelligent Virtual Assistants. *Computers in Human Behavior*, 135, 107362. <https://doi.org/10.1016/j.chb.2022.107362>

Bräuer, P., & Mazarakis, A. (2022). How to Design Audio-Gamification for Language Learning with Amazon Alexa? - A Long-Term Field Experiment. *International Journal of Human-Computer Interaction*. <https://doi.org/10.1080/10447318.2022.2160228>

Bräuer, P., Werner, F., & Mazarakis, A. (2023). Extracting Game Design Elements from Voice-Enabled Games: A Review of Amazon Alexa Skills. *Proceedings of the 7th International GamiFIN Conference*, 47–56. **(received the best paper award)**

In addition, a complete list of all of the author’s publications produced during candidacy can be found in the appendix.

1.2. Research Questions

The aim of this thesis is to empirically investigate the effect of purely acoustic gamification in the context of IVAs. The idea has been developed out of the field of gamification research. Figure 1 summarizes how the idea was developed based on the concept of gamification, which is introduced in section 2.1.

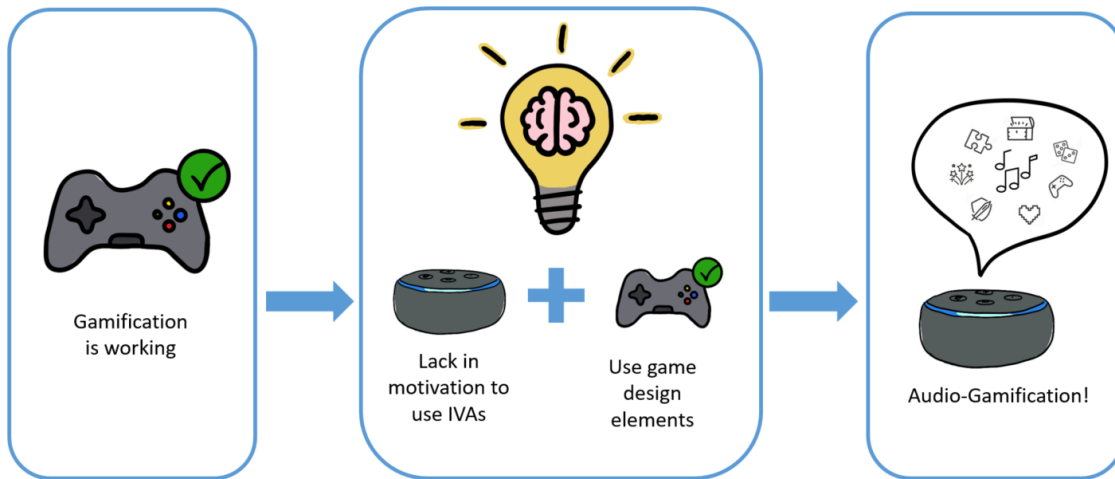


Figure 1 Process of developing the research idea and questions

Other studies with IVAs have shown that this new technology has a problem with declining use over time (Cho et al., 2019; Lima et al., 2023). Users buy a smart speaker or try out the voice assistant on their smartphone, but the initial interest fades, and voice control is only used occasionally to play music. However, there are benefits to voice interaction that make it worthwhile to support the adoption and continuous use of this technology. Gamification has already proven its value in supporting the diffusion of technologies, as shown by the example of fitness trackers (Vooris et al., 2019). These devices show the user how many steps they have taken while walking or running. Gamification features have proven to be crucial for the adaptation of fitness trackers (Spil et al., 2017). Primarily by connecting to an app, fitness trackers allow users to compete with themselves or others through challenges. Unlocking achievements is also one of the various gamification approaches used by manufacturers of these devices to make physical activity more motivating. By setting goals, users are encouraged to move a little further (Vooris et al., 2019). Gamification elements also seem crucial for the continued use of fitness trackers (Windasari & Lin, 2021). The example shows how gamification can be used to convince users to adopt a new technology in their daily lives. However, the combination of gamification and IVAs is not simply a new example of gamification implementation due to the limitations of audio channels. Rather, it represents a new subconcept.

To clearly define this concept, the term audio-gamification is used. According to this definition, audio-gamification is the use of game design elements in a non-game context, which are presented only via audio output (Bräuer & Mazarakis, 2022a; Mazarakis, 2021). From a scientific point of view, whether this concept of audio-gamification can be

implemented has not yet been considered. The overarching research gap this thesis addresses is summarised in the following question:

Research Gap: *How does audio-gamification impact user motivation and performance in the context of Intelligent Virtual Assistants?*

Four research questions are addressed in this thesis to fill this research gap. Each of these four research questions stands on its own and can be answered individually. In terms of the research gap addressed here, the individual research questions and their answers fit together like puzzle pieces that come together to form a picture. The relationship between the research questions and how they build on each other is shown in Figure 2.

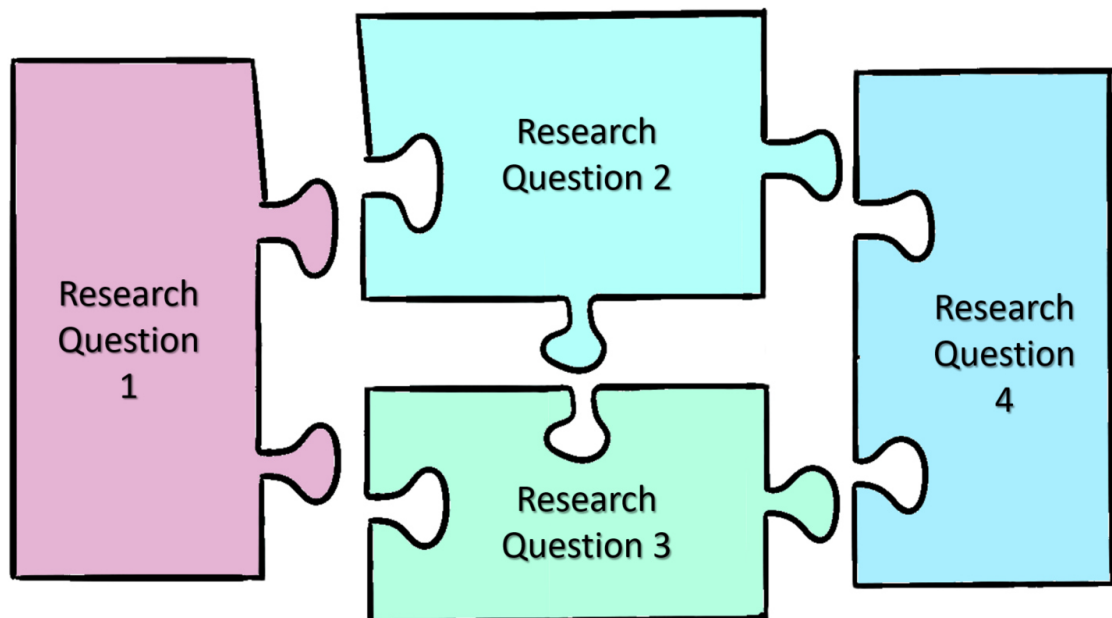


Figure 2 Relationship of the four research questions to address the research gap.

The first research question deals with gamification on the level of individual game design elements and thus creates the basis for the following investigations. Numerous studies have demonstrated the effectiveness of gamification in various contexts (Dichev & Dicheva, 2017; Hamari et al., 2014). However, many examples of gamified applications rely on relatively extensive concepts that combine many different game design elements (Domínguez et al., 2013; Sailer et al., 2013). This multitude of game design elements is usually visually integrated into the gamified application. However, if one wants to convey information solely through audio, as with an IVA, one can quickly overwhelm users by providing too much information simultaneously (Garcia et al., 2013). So, is it enough to pursue very simple approaches to gamification by merely relying on individual game

design elements? As a basis for further research on gamification in the context of IVAs, the following research question is therefore posed first:

Research Question 1: *How do different game design elements affect user motivation?*

Following on from the aspect of the individual game design elements, the focus is placed on the general feasibility of audio-gamification. The assumption is that gamified approaches can be implemented without a visual component. This assumption is based on the considerations mentioned in the introduction and different concepts of video games that work without complex graphics (these will be presented in section 2.3.). A research question was formulated to investigate the effect on motivation and performance, as it could be achieved in other studies (Bicen & Kocakoyun, 2018; Putz et al., 2020; Sailer et al., 2017) by visually implemented game design elements:

Research Question 2: *If and how does audio-gamification affect user motivation and performance?*

One aspect that is repeatedly questioned is the external validity of studies in gamification (Landers et al., 2018) and IVA research (McTear et al., 2016). Many results are collected in laboratory settings with relatively small samples that can be well-controlled but do not necessarily reflect real-world subject behavior (McTear et al., 2016). The effectiveness of gamification can be observed across several popular platforms such as Duolingo, and Uber, but most studies are set up in labs and conducted with students. Therefore, another research question can be completed by examining the effect of audio-gamification on users in a natural environment. As shown in Figure 2, the third research question builds on research question 1 and also connects to research question 2. In accordance with research question 1, the focus is also on examining individual game design elements while simultaneously applying them to the context of audio-gamification as per research question 2.

Research Question 3: *How do different game design elements in the context of IVAs affect motivation and performance of real-world subject behavior?*

Since interaction via voice and without a screen has different design requirements, implementing game design elements via audio is challenging. Gamification concepts that

have been proven in research and practice, such as the introduction of badges or leaderboards (Koivisto & Hamari, 2019), do not seem to be easily transferable to audio-gamification. In order to develop concepts for audio-gamification that are as suitable as possible, it is necessary to investigate which gamification options are particularly well suited for the audio context. Existing concepts are usually studied in one of two ways, either by analyzing the literature or existing applications. A literature review is not feasible as this thesis is pioneering in dealing with audio-gamification. However, examining games from which game design elements can be extracted is appropriate since there is a large selection of such games on the various IVA platforms. Finally, the following research question is posed:

Research Question 4: *What are the dominant game design elements used in voice-enabled games for Amazon Alexa?*

This last research question ties into RQ2 and RQ3 (see Figure 2), in which the findings on the impact of audio-gamification are extended by a set of auditory implementable game design elements. This establishes a foundation from which practical implementation concepts and further research can be derived. In summary, answering the four research questions will help answer the overarching main research question. The following chapters present the individual studies conducted as part of this thesis to answer the research questions.

1.3. Outline

| | |
|---|--|
| Part 1 INTRODUCTION & FOUNDATION | Chapter 1 Introduction |
| | Chapter 2 Related Work |
| Part 2 THE STUDIES | Chapter 3 Examination of Individual Game Design Elements (RQ1) |
| | Chapter 4 Audio-Gamification Lab Experiment (RQ2) |
| | Chapter 5 Audio-Gamification Field Experiment (RQ1 & RQ3) |
| | Chapter 6 Audio Game Design Element Extraction (RQ4) |
| Part 3 CONSOLIDATION OF RESULTS | Chapter 7 Discussion of all Studies |
| | Chapter 8 Conclusion |

Figure 3 Structure of this Thesis

Chapter 2 begins with a review of related work and covers the current state of research in the areas of gamification and intelligent virtual assistants. The introduction to both topics demonstrates the existing links and identifies possible challenges.

Chapter 3 comprises the first publication in which the results of a field study are presented. The study addresses the effect of individual game design elements on motivation. It serves as a starting point for employing straightforward gamification techniques rather than complex multilayered systems. These initial results form the basis for the idea of audio-gamification, demonstrating that even the inclusion of a single game design element can increase motivation.

Chapter 4 presents the results of a laboratory experiment. The goal of this study was to investigate whether the use of audio-gamification can make an application for Amazon Alexa more motivating. For this purpose, a skill for Amazon Alexa was enhanced by a series of acoustically implemented game design elements compared to a control group. It is the first study of audio-gamification and is an indication of the immense potential of IVA design.

The results of a long-term field study that builds on the findings of the two previous studies are presented in *Chapter 5*. An Amazon Alexa skill for foreign language learning was developed for the field study. Two different variants of audio-gamification were implemented and compared to a control group. The study demonstrates the potential of audio-gamification for language learning in a non-artificial setting, producing very valid results.

The last study of the thesis is presented in *Chapter 6*. Unlike the other three studies, this one is not based on an experiment but on an analysis of Amazon Alexa games. The games were dissected into their game design elements to gain insight into which elements are used by IVA developers and how they can potentially be transferred to gamify IVA applications.

Chapter 7 discusses and relates the results of all four studies. In addition, an outlook is given on how future research work could follow from this thesis. In *Chapter 8*, the results of this thesis are summarized.

CHAPTER

TWO

Related Work

In this chapter, the underlying topics of gamification and IVAs are introduced to approach the topic of audio-gamification. Furthermore, the chapter also briefly explores the field of video games that function without graphics.

2.1. Gamification

The term gamification was first introduced in 2002 (Pelling, 2011), but it was not defined more precisely at that time. Various authors have since tried to grasp the term and have created partly very different definitions, focusing either on game design elements or game mechanics or the process of gaming or gameful experiences in non-gaming contexts (Krath et al., 2021). Since gamification is often applied in marketing, Huotari and Hamari (2017) created a definition explicitly adapted to this context. The authors understand gamification as “a process of enhancing a service with affordances for gameful experiences in order to support users’ overall value creation.” Zichermann and Cunningham (2011) also highlight the process of gamification in their definition and refer to gamification as “the process of game-thinking and game mechanics to engage and solve problems.” However, the definition that has become widely accepted was established by Deterding et al. (2011). This definition defines gamification as the use of game design elements in a non-game context. In the context of this thesis, this definition, according to Deterding et al. (2011), is used as a basis since it describes the concept more generally and does not limit it to a specific context.

Gamification is applied in various contexts, such as business (Wunderlich et al., 2020), sports, and education (Dichev & Dicheva, 2017; Sailer & Homner, 2020). It has now

become a well-established research area in human-computer interaction (Rapp et al., 2019). There are many examples of everyday applications that use gamification: Fitness trackers that count the user's steps and award badges for reaching certain step counts or learning apps (Shortt et al., 2023) that give users daily challenges and allow them to compare themselves to other learners on a leaderboard.

To provide a clear definition of gamification, it may also be helpful to consider what it is not. First, it is crucial to distinguish game from play. According to Caillois (2001), playing (*paidia*) is a free form of play in which improvised, unstructured behavior can be observed. In contrast, gaming (*ludus*) describes a rule-based game with clear goals and quantifiable results. Gamification refers to the rule-based form of gaming. In other words, it's more about achieving specific goals and less about playing freely and creatively according to your own ideas.

Although gamification builds on the foundations of rule-based gaming, gamification should not be misunderstood as a game (Mazarakis, 2021; Palmquist, 2021). As Deterding et al. (2011) describe, it is about using elements usually used in games to increase the motivation or performance of users in other contexts. The primary goal is not to entertain the user but to create incentives to perform certain activities. An adjacent concept is so-called serious games. These are full-fledged games that, like gamification, serve purposes other than pure entertainment (Susi et al., 2007). Serious games are used to teach skills and knowledge, simulate processes (Susi et al., 2007), or address social and ethical issues (Pereira et al., 2012). But unlike gamification, serious games are a complete standalone application. Gamification, on the other hand, is a tool that is used in a complementary way and does not work without the context in which it is used (e.g., a learning app) (Mazarakis, 2021).

Another distinction that should be made when considering gamification is the concept of rewarding. A successful gamification concept does not require physical or monetary rewards. Compensating people for an activity, such as giving them a salary for their job or introducing a penalty for breaking the rules, usually works initially. Still, the effect wears off, and the reward is perceived as insufficient (Werbach & Hunter, 2012). For example, a study by Xu and Hamari (2022) has shown that gamification can promote creativity more strongly than monetary incentives. A classic example of rewarding is customer loyalty programs such as Miles and More or Payback (Reinartz, 2010). In these programs, customers receive points for their purchases, which they collect and can later

exchange for rewards. Customers are not primarily motivated by collecting the points but rather by the possibility of using the points as currency to exchange them for real value.

Gamification, on the other hand, does not rely on this form of reward but motivates users solely through the effect of the game design elements. How exactly this motivating effect is created can be explained by different theoretical approaches. In the next chapter, three motivational theories will be presented.

2.1.1. Motivation Theories

The goal of using gamification is to motivate users (Warmelink et al., 2020). Several theoretical concepts describe how a motivating effect can be generated (Zainuddin et al., 2020). In this chapter, three theories are presented that are frequently used to describe motivational effects: the self-determination theory, the flow theory, and the goal-setting theory.

In gamification research, the most commonly referenced is the so-called Self-Determination Theory (SDT) (Seaborn & Fels, 2015) by Deci and Ryan (2000). According to SDT, there are two types of motivation. Either an individual can be extrinsically motivated, i.e., motivated by incentives that come from outside, or intrinsically motivated, i.e., through an activity of one's own volition. For an individual to be intrinsically motivated, SDT assumes that three so-called basic psychological needs must be met (Deci & Ryan, 2000). The first of these basic needs is the experience of competence; it relates to a need for challenge and feelings of effectance. The second basic need describes a sense of volition or willingness when doing a task and is referred to as autonomy. The third basic need is social relatedness, which occurs when someone feels connected to others (Deci & Ryan, 2000; La Guardia et al., 2000). In gamification research, SDT is used to derive design implications and measure the effects of interventions in experiments (Krath et al., 2021).

Another theory often used in gamification research is the flow theory (Csikszentmihalyi, 2009). Csikszentmihalyi describes a state workers can reach when they are entirely absorbed in their work. This state of flow is characterized by intense concentration, a fusion of action and consciousness, loss of self-consciousness, and distortion of temporal experience. This phenomenon can also be observed when playing video games when the player is in a "tunnel" and wholly absorbed in the game. For gamification, flow theory

can describe the balancing of an application for optimal user experience (Marczewski, 2018) primarily since flow is supported by clear goals and feedback (Csikszentmihalyi, 2014), which are also of great importance for gamification. However, measuring flow is challenging, and previous results from gamification research are inconsistent (Oliveira et al., 2021). This raises doubts about the feasibility of using the theory in experimental studies within the field. Therefore, there has been no in-depth study of flow in this thesis.

Gamification research increasingly applies goal-setting theory as a basis (Zainuddin et al., 2020). This theory is based on the observation that difficult goals produce higher levels of performance than easy goals, and that specific difficult goals produce higher levels of performance than ambiguous difficult goals (Locke, 1968). The approach assumes that a mediator effect between setting goals and performance occurs through the psychological process of self-regulation. Self-regulation can be defined as the modification of thoughts, affects, and behavior (Karoly, 1993). To increase motivation, goal setting is a very effective method (Locke & Latham, 2002). For example, Landers et al. (2017) apply the theory as the basis of an experiment to show how leaderboards can provide goals to increase performance. According to the study, the mere existence of a leaderboard motivates users to strive for the top spot, even if it is not explicitly stated as a goal. Also, other game design elements, such as achievements, have already been studied in the context of goal-setting theory and have shown that this theory provides a suitable basis for explaining motivational effects through gamification (Groening & Binnewies, 2019). The impact of gamification and how it can be embedded in different theoretical motivational frameworks depends, among other things, on which game design elements are considered. The next chapter introduces the topic of game design elements and presents different frameworks that give an overview of some of the more popular game design elements.

2.1.2. Game Design Elements

In the scientific context, game design elements play a crucial role when talking about gamification. The definition of gamification, according to Deterding et al. (2011), focuses on the game design elements and clarifies how relevant they are for gamification. The authors describe game design elements as characteristic of games, found in many games (though not necessarily all), associated with games, and playing a significant role in games. More recently, they have been referred to as building blocks that are

characteristic of games or gamified systems (Fortes Tondello, 2019). However, game design elements do not include all elements found in games. For example, the game menu is not included; although most video games have a menu, it is usually irrelevant to the gameplay experience (Tobar, 2017).

Various publications have attempted to create frameworks that are as comprehensive as possible and provide an overview of the selection of elements that can be used to gamify applications in various contexts (Toda et al., 2019). In the different frameworks, game design elements are mapped in various degrees of detail and classified into groups according to different approaches. In the following, two widely-used frameworks are presented (Mora et al., 2015): one outlined by Werbach and Hunter, and the other by Marczewski.

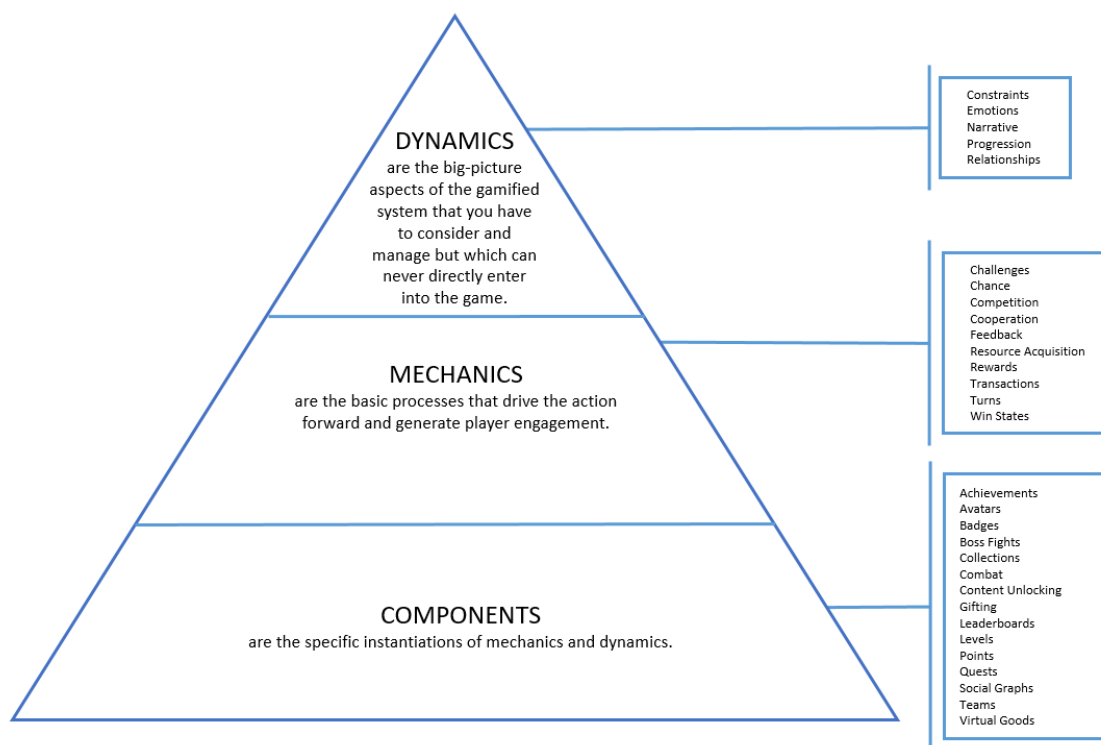


Figure 4 Model of game design elements by Werbach and Hunter adapted from (Werbach & Hunter, 2012)

Werbach and Hunter (2012) present a three-level model based on Hunicke et al.'s (2004) Mechanics, Dynamics, and Aesthetics Model (MDA) from game design. Werbach and Hunter's model contains a total of 30 elements that are distributed over three levels. The illustration of the model and its components is shown in Figure 4. On the lowest level are the so-called components, above them are the mechanics, and on the top level are the dynamics. The dynamics include five abstract elements, such as emotions and conditions, and are manifested by the mechanics. The ten mechanics are meant to represent

fundamental processes that drive user actions and thus create engagement. Examples of mechanics include challenge, feedback, collaboration, and competition; a complete overview can be found in Figure 4. Most specific are the fifteen components that specify the mechanics. However, the model does not describe how the elements on the different levels of the model are connected. The lack of connections leaves the developer somewhat on their own when it comes to selecting suitable elements.

Another relatively comprehensive collection of typical elements that can be used to gamify an application is presented by Marczewski (2018) in his periodic table of gamification elements. His periodic table contains a total of 52 elements, which the author divides into eight different player profiles. The author assumes that the different game design elements support the different player profiles differently. These player profiles correspond to Bartle's Hedax model (Bartle, 1996), which assumes six player types: Achiever, Player, Socialiser, Philanthropist, Free Spirit, and Disruptor. Based on this concept of player types, a well-balanced gamification concept can be developed from the set of elements that fit the respective target group and the respective context.

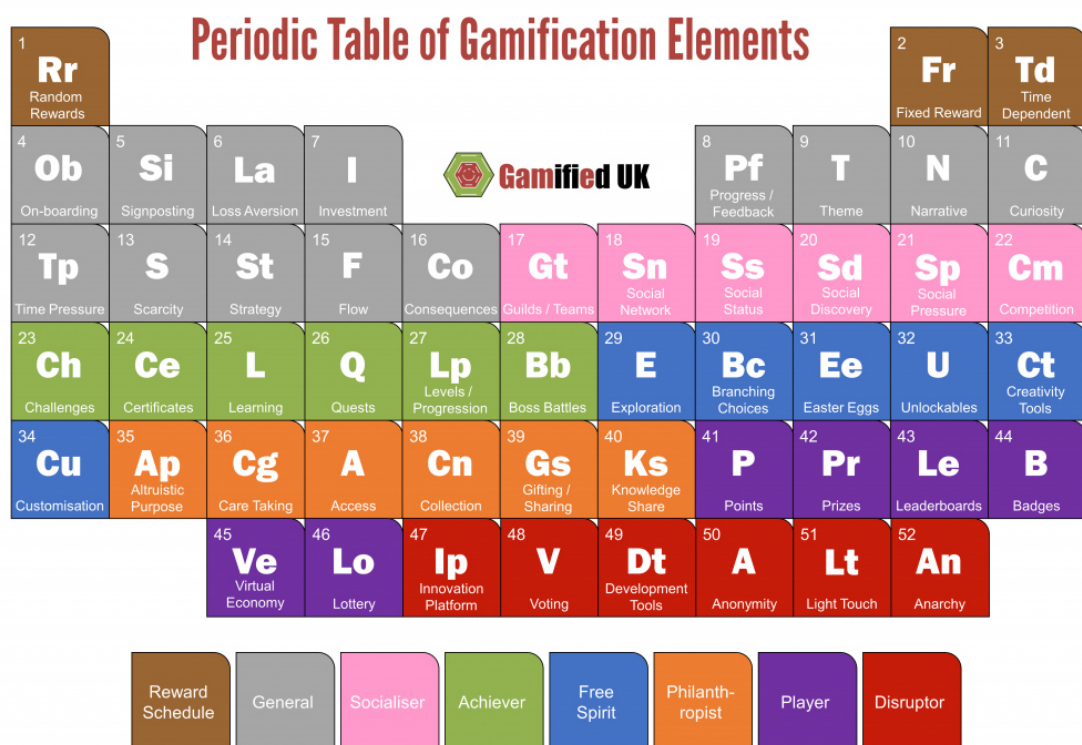


Figure 5 The Periodic Table of Gamification Elements by (Marczewski, 2018)

In terms of game design elements, there is an overlap between the two models despite the different approaches to the underpinnings of the elements. Furthermore, both models

share the basic assumption that different game design elements can be used for different purposes. It has also been observed that game design elements can affect motivation differently (Mekler et al., 2017). Although gamification uses only game design elements and does not require building a game, implementation is not easy (Kapp, 2012). A gamification concept must be adapted to the context in which it will be applied. It is not enough to simply integrate classic game design elements such as points, badges, and leaderboards into an application. For example, the awarding of points should be adapted to the tasks (Mazarakis & Bräuer, 2022), a story should fit the context of the application (Mazarakis, 2017), and the introduction of a leaderboard should take into account that new users do not necessarily want to compare themselves with long-term users (Mazarakis, 2015). Exactly how the different game design elements work, whether individually or in various combinations, is still a rather broad and open area of research that will be explored in more detail in chapter 3.

2.2. Intelligent Virtual Assistants

Intelligent Virtual Assistants (IVAs), like Amazon Alexa, Google Assistant, and Siri, are now available on a wide variety of devices, in smartphones, at home on smart speakers, in cars (Cowan et al., 2017), or via wearables such as smartwatches and glasses (McTear et al., 2016). IVAs belong to the Conversational User Interfaces (CUI) category, which includes all software systems that enable interaction between humans and machines via natural language (McTear et al., 2016). IVAs are characterized by the capability of interaction through spoken language. This distinguishes them from chatbots, usually operated by text input via a keyboard (Rzepka et al., 2022).



Figure 6 IVAs are used especially on so-called smart speakers. Shown are an Amazon Echo Dot 4th generation (left), 3rd generation (center), and an Amazon Echo Show (right).

In addition to the term IVA, there are several other terms in the literature that refer to the same concept: Intelligent Personal Assistant (IPA), Conversational Agents, Virtual Personal Assistants, Personal Digital Assistants, Voice-Enabled Assistants, or Voice Activated Personal Assistants (Cowan et al., 2017). Currently, IVA and IPA appear to be widely accepted. In the following, the term IVA will be used.

2.2.1. Benefits of IVAs

There are several reasons why interacting with a machine via voice is advantageous. Two advantages that stand out, in particular, are that IVAs enable touch-free and eyes-free interactions (Bouزيد & Ma, 2022; McTear et al., 2016), which can significantly benefit various situations. For instance, while operating a car, voice commands allow the driver to control the music player or navigation system without needing to look away from the road. While cooking, one can check the next steps of a recipe without the need to clean one's hands to avoid smudging the smartphone. Order picking in logistics is a further example. One can query the location of a specific product using voice commands and have hands free to pick the items.

The use of voice input and audio output for queries requires minimal effort, particularly when the user has a simple request, such as checking the weather for the next day (Bouزيد & Ma, 2022). If the same question is to be answered via a smartphone app, for example, the user must first unlock the device, then click on the correct app and select the correct

city and day in the app. The query via IVA, on the other hand, is completed with a spoken sentence and is, therefore, significantly faster. Overall, entering texts is much faster via speech than typing (Ruan et al., 2018).

Communication via voice also has the benefit of not requiring the user to be able to read or write in order to engage with the machine (Bouزيد & Ma, 2022). Thus, an IVA can be operated by children (Sakr, 2021) or illiterate individuals (Da Silva et al., 2022) intuitively. Similarly, IVAs offer great potential for individuals with visual impairments (Abdolrahmani et al., 2018; Pradhan et al., 2018).

Meanwhile, various smart speakers are also sold with a screen (see Figure 6 right Amazon Echo Show), and our smartphone's IVA can visually display information on the device. Although a screen can alleviate some of the difficulties in transmitting information by voice alone, this technology loses its appeal. This implies that the user needs to be in close proximity to the device and direct their attention and gaze toward it. The main advantage of this technology is its lack of a display. However, designing IVA applications requires a shift in thinking. The subsequent section addresses the challenges that come with this shift. However, before delving into these challenges, a brief overview of the technical implementation of IVAs will be provided since many challenges arise from technical nuances.

2.2.2. Technical Implementation of IVAs

The development of IVAs has been driven by advances in artificial intelligence (AI) and machine learning (ML), which are prerequisites for speech processing and understanding in these systems (McTear et al., 2016). To respond to a request at any time, IVAs continuously listen and capture what is being spoken in their environment. However, according to the manufacturers, a user's utterance is only accepted and processed if a so-called wake word is recognized (Pearl, 2016), such as "Alexa" or "OK Google."

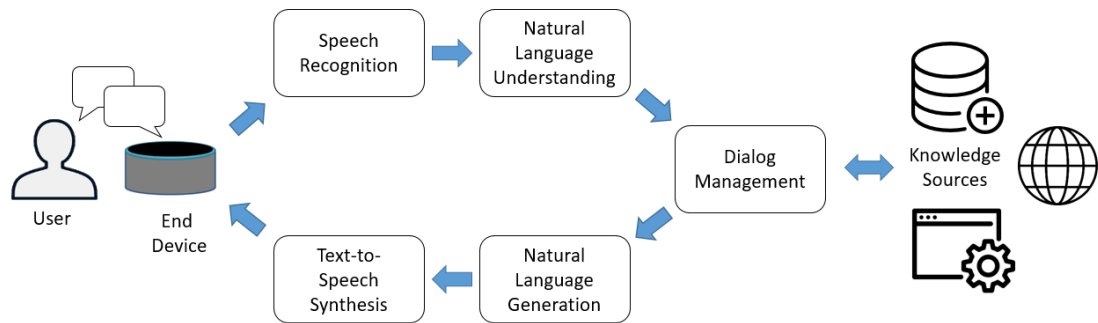


Figure 7 Generalized Dialogue Architecture (following M. McTear et al., 2016)

Figure 7 illustrates the process of dialog between the user and the IVA. The system receives input in spoken language, which is converted into a word string using speech recognition. The word string is interpreted using natural language understanding to determine the meaning of the user input. Based on this interpretation, the system's dialog manager decides how to respond to the user's input. A response can take various forms, for example, an answer to a search query, launching an application, or generating a follow-up question if the query is incomplete. The required information can be obtained and processed to generate a response from various knowledge sources. For example, knowledge sources can be databases explicitly created for the application or APIs for other web services. The generated response is then used to create a response string via natural language generation, from which audio output is generated via text-to-speech synthesis (McTear et al., 2016).

Interacting with a machine via voice input can provide very different degrees of flexibility in input (McTear, 2017). The user can be provided with predefined options, similar to buttons in a Graphical User Interface (GUI), whose actions generate specific responses. However, CUIs can also handle more complex queries, such as variations in wording by the user and adaption to context over the course of a conversation, depending on the complexity of the system logic (Bouزيد & Ma, 2022).

In order to expand the range of functions an IVA can fulfill, the manufacturers (e.g., Google and Amazon) offer the possibility of enabling third-party applications via their platforms. These applications can be compared to smartphone apps and are called *skills* for Amazon Alexa (Hoy, 2018). A user-defined interaction model is created to develop a skill that requires logic and a voice user interface (VUI). A VUI allows the user to interact with the skill by defining what actions should be performed when the user makes a specific *utterance*. The skill therefore extends the basic vocabulary of utterances that an

IVA can process. Depending on the context in which an utterance is made, it can be based on different intentions. For example, the utterance “Yes next” may be in response to different questions such as “Should I start the next song from the playlist?” or “Do you agree with the terms of the contract?” During the development of the voice user interface, it is defined which intention the user is likely to have with certain utterances. These intentions are referred to as *intent* (McTear et al., 2016).

Each intent is assigned a specific action. In other words, an intent correlates to a function name that may be invoked in programming logic. A developer defines a set of possible utterances a user could use to express a particular intent. Using these example utterances in combination with ML, the system can then infer which intent is meant by the user with a specific utterance, even if the user’s wording does not precisely match the given examples (McTear, 2021). An example of an utterance that triggers an intent `check_weather` could look like the second utterance of the user in the following dialog.

```
User: "Alexa, start weather broadcast."  
Alexa: "Welcome to weather broadcast. How can I help you?"  
User: "Tell me, how is the weather in Kiel today?"
```

For more complex user requests, utterances can also contain variables that help to make an input easier to interpret. These variables are called *slots* (or entities) (Edo & Lucio, 2021; McTear et al., 2016). In the example above, such a slot could be the city for which the weather is requested or the day. The user could, therefore, also ask for the weather in Hamburg for the day after tomorrow, but not every utterance would have to be stored explicitly. To illustrate this, in the following example, the city of Kiel is replaced by the slot `citySlot` and the day by the slot `daySlot` in the user request. The `citySlot` contains a list of possible cities that could appear in a request for the weather.

```
User: "Tell me, how is the weather in {citySlot} {daySlot}?"
```

An *invocation name* must be defined before a user may begin using a skill. To start the skill, the user then only needs to name the wake word and the invocation name. In this example, the invocation name would be “weather broadcast” and used in the user’s first utterance combined with the wake word “Alexa.”

In addition to the voice interface, a logic is created for each skill. This corresponds to a web service containing the functions executed when a specific intent is called. The skill logic is hosted on a separate server and communicates with the VUI via HTTPS (*Host a*

Custom Skill as a Web Service | *Alexa Skills Kit*, 2022). Other CUI's, such as chatbots, can be realized using the same principle.

2.2.3. Introduction to Alexa Skill Development

Building on the structure of IVA applications described in chapter 2.2.2., this chapter provides insights into the development of skills for the Amazon Alexa platform. There are various frameworks for implementing skills whose functions are all relatively similar and only differ in syntax. The following examples describe some basic steps for implementation using the JOVO framework in version 3. This open-source framework allows the development of voice-based applications that are compatible with several platforms at the same time, e.g., Amazon Alexa and Google Assistant. This framework was used to implement the skill in Study 3, designed for a productive implementation for the Amazon Alexa Store and models more complex dialogs. The examples used here are simplified excerpts from the implementation of the skill. The context of the skill is an application for learning German as a foreign language. The skill asks for vocabulary either in one of two gamified variants or without gamification.

In addition to the developed source code, the basic requirement for the productive operation of a skill on the Amazon Alexa platform is the setup of the skill backend on the hosting platform. The backend consists of two parts: the Alexa developer console and the endpoint on which the skill's logic is hosted. The communication between these two parts must be configured in both places, which is not described here but should be taken from the respective documentation.

The links between the functions of the source code and the intents, slots, and invocation name must be defined on the developer console. These links are defined using a language model (also called interaction model). All utterances and slots are stored in the language model. The following example shows an extract from the model of the skill from Study 3. First, the invocation name of the skill is defined in the model. Then, the name and the associated slots are stored for each intent, called inputs in the JOVO model (v3). Each input has a name and a type. The type corresponds to a slot type, which specifies how data from the slot is recognized and handled¹.

¹ <https://developer.amazon.com/en-US/docs/alexa/custom-skills/slot-type-reference.html>

Amazon provides predefined slot types. These predefined slots recognize certain sets of input and do not need to be added manually². In the example below, the predefined slot AMAZON.FirstName was added, which contains a large set of first names and nicknames. In the skill, this slot type is used when the user is asked for the name by which they are greeted and which is used in a gamified variant for a leaderboard. After assigning slot types to Intents, the slots for the slot types are defined at the end of the model. Each slot type name is assigned different slots (labeled as value in the example). In the example, there is a slot type “answer“, which in this abbreviated example contains two words. These correspond to the possible correct answers a user can give when the skill asks for a German word. Possible synonyms are also stored for each slot. For example, the answer to the question "What means Flugzeug in English?" is both airplane and plane. The complete model can be stored as a JONS file in the Amazon Alexa Developer environment.

```
{
  "invocation": "german vocabulary coach",
  "intents": [
    {
      "name": "GameWelcomeContdIntent",
      "inputs": [
        {
          "name": "name",
          "type": {
            "alexa": "AMAZON.FirstName"
          }
        }
      ]
    },
    {
      "name": "AnswerIntent",
      "inputs": [
        {
          "name": "answer",
          "type": "answer"
        }
      ]
    }
  ],
  "inputTypes": [
    {
      "name": "answer",
      "values": [
        {
          "value": "to sail",
          "synonyms": [
            "sale",
            "sail"
          ]
        }
      ]
    }
  ],
}
```

² <https://docs.aws.amazon.com/lex/latest/dg/howitworks-builtins-slots.html>

```

        {
            "value": "airplane",
            "synonyms": [
                "plane"
            ]
        }
    ]
}

```

The logic of each intent must be implemented to match the language model. As with slots, some intents come predefined. Both Amazon and the various frameworks provide some of these predefined intents and slots. In the JOVO framework, these include the `LAUNCH` and `END` intents. When a skill is started by calling the invocation name, the `LAUNCH` intent is executed first. The `END` intent is executed to terminate a session. The intent can be triggered in several ways:

- either when the user uses intents that are directly associated with the intent, such as `Stop` or `End`;
- when an error occurs;
- when the skill times out because the user does not respond;
- or when another intent triggers the `END` intent.

The following example uses both intents and adds a simple dialog with custom intents.

```

app.setHandler({
  LAUNCH() {
    // Ask a yes-no-question and route to next state
    this.$speech.addText('Do you want to start learning?');
    this.$reprompt.addText('Please answer with yes or no.');
```

```

    this.followUpState('StartState')
      .ask(this.$speech, this.$reprompt);
  },

  // Behave differently for a 'yes' or 'no' answer inside answer state
  StartState: {
    YesStartIntent() {
      // Ask a yes-no-question and route to next state
      this.$speech.addText('Great, would you like to pick up where
        we left off last time?');
      this.$reprompt.addText('Please answer with yes or no.');
```

```

      this.followUpState('AnswerState')
        .ask(this.$speech, this.$reprompt);
    },
  },
});

```

```

        NoStartIntent() {
            // Do something
        },
    },

    AnswerState: {

        YesAnswerIntent() {
            // Do something
        },

        NoAnswerIntent() {
            // Do something
        },
    },

    END() {
        // Do something
    }

});

```

Before the first custom intent `YesStartIntent()` is defined, it is assigned to a state. States are needed for more complex dialogs to assign a user input to the correct context. In the example above, the user can answer both the question “Would you like to start learning?” and the question “Would you like to continue where we left off?” with “yes” or “no.” Depending on the state of the dialog, the user's answer will have a different result. In the example above, if the user answers “yes” twice, the user is taken from the `LAUNCH` intent to the `StartState`, and from there to the `AnswerState`.

When an intent is to be passed to another state, `followUpState` is called. In addition to the state's name to follow, a question and a reprompt are passed to the `ask` method. The `ask` method activates the microphone and expects input from the user. If the user does not answer the question, the reprompt is played³.

```

this.followUpState(stateName)
    .ask(speech, reprompt);

```

In the example above, the user is asked whether they want to start learning. Regardless of whether the answer is yes or no, the user is always forwarded to the `StartState`. The answer, whether “yes” or “no,” then determines which intent is executed within this state.

In addition to processing user input for calling intents, the input can be processed in other ways. This is done by accessing the `$inputs` object⁴. The following example checks

³ <https://v3.jovo.tech/docs/output#ask>

⁴ <https://v3.jovo.tech/docs/routing/input>

whether the value of the input, which is of slot type answer, matches the name of the first value from the array userWords, i.e., the word that was just asked for. If this is the case, the answer is correct, and the user receives feedback accordingly.

```
NonGameAnswerState: {
  // check input for correctness
  async AnswerIntent() {

    // compare the input with the list of vocabulary words
    if (
      await oneOf(
        this.$inputs.answer.value.toLowerCase(),
        userWords[0].name,
        this,
      )
    ) {
      //give feedback gamified or not, depending on group
      if (this.$session.$data.isGamificated == 0){
        await answerNonGamified(this, true, true)
      }
      else {
        await answerGamified(this, true, true)
      }
    }
    // answer was not correct, give negative feedback
    else {

      // do something
    }
  }
}
```

If a skill is to be operated productively, i.e., it can be used by several users online in parallel, information must be stored in the backend about which user is currently at which point in the dialog. This information is stored via session data. A session consists of at least one request and one response and represents an uninterrupted dialog between the user and the skill⁵. The concept of a session is illustrated in Figure 8.

⁵ <https://v3.jovo.tech/docs/requests-responses>

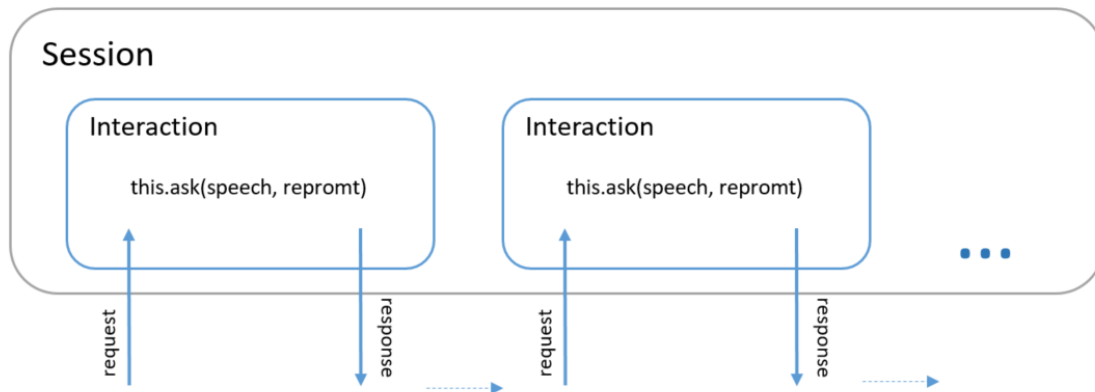


Figure 8. Illustration of the concept of sessions based on the JOVO documentation

Various information can be stored in elements for sessions. The structure is as follows:

```
this.$session.$data.key = value;
```

In the above example, the session contains an `isGamificated` element that specifies whether or not the dialog of the current session is gamified. If the query `this.$session.$data.isGamificated == 0` is true, a non-gamified response is returned.

Session data can be stored and retrieved in a database to store user data across multiple sessions. This is important, for example, if you want to document a user's progress to gamify a system. Amazon Alexa is typically combined with DynamoDB, a NoSQL database provided by AWS. However, other databases can be integrated as well. The following example integrates a MySQL database⁶.

```
const { MySQL } = require('jovo-db-mysql');

// Enable DB after app initialization
app.use(
  new MySQL()
);

// Establish MySQL connection
var mysql = require('mysql');
// MySQL connection parameters, need to be changed for production
var con = mysql.createConnection({
  host: 'localhost',
  user: 'root',
  database: 'alexa_skill'
});

con.connect(function(err) {
  if (err) throw err;
});
```

⁶ <https://v3.jovo.tech/docs/databases#docs>

```

app.setHandler({
  async LAUNCH() {
    // If user starts skill for the first time, store required data
    // inside the database and randomly decide if to gamify or not
    if (this.$user.isNew() && !await hasEntry(this)) {
      this.$session.$data.isGamified = Math.floor(Math.random() *3);
      this.$session.$data.insertUser = "INSERT INTO user (id, gamified,
        points) VALUES (" + mysql.escape(this.$user.getId())
        + "," + mysql.escape(this.$session.$data.isGamified) + ","
        + mysql.escape(0) + ")";
      await useDB(this.$session.$data.insertUser);
    } else
      //do something
  }
})

async function hasEntry(object) {
  // construct some database query
}

// function for every query to the database
function useDB(query) {
  // do a database query
}

```

The code snippet first enables MySQL integration. Then, the parameters for connecting to the database are defined. When the skill is started, the LAUNCH intent uses the `isNew()` function provided by JOVO to check if the user is known and the `hasEntry()` function to check if there is no entry for the user ID in the database. If both conditions are met, a random value between 0 and 2 is assigned to the `isGamified` element in the session data, which determines the version of the dialog that will later be presented to the user. The string for the database query to create a new entry in the user table is then written to the `insertUser` element. In this abbreviated example, values are passed for the three columns: `id`, `gamified`, and `points`. Finally, the string is passed to the `useDB` function that executes the database query.

2.2.4. Challenges in The Design of IVA Applications

Despite the advantages of communicating through speech, the design of speech-based applications is not straightforward. This is also reflected in the repeatedly observed decline in use (Cho et al., 2019; Lima et al., 2023; Lopatovska et al., 2019; Luger & Sellen, 2016). Many users have expectations of an IVA that have not yet been met (Cho et al., 2019). Users expect a natural conversation, as they would have with another person. However, the responses of an IVA usually do not vary or vary only slightly, which becomes boring and irritating over time. Another problem that leads to limited use is the

precision of the requests and the lack of visual feedback (Luger & Sellen, 2016). This becomes relevant, for example, when users want to send longer messages. If the IVA gets something wrong, it is difficult to detect and correct. Users may also be frustrated by incorrect or incomplete responses (Lopatovska et al., 2019). Lack or loss of trust can also be a barrier to regular use (Luger & Sellen, 2016). For tasks such as buying a ticket to an event on a specific date where the user is spending money, even a misunderstanding can lead to a feature not being used again. A lack of empathy and personalization can also lead to a rejection of the assistant. For example, a user may tell the assistant about a problem at work and receive only general advice in response (Cho et al., 2019). There are several approaches to addressing these issues in developing skills for IVAs, which are presented now.

Designing a dialog for a VUI differs significantly from writing a text or developing a classic GUI (Murad & Munteanu, 2020). When explaining something, the phrases used may differ depending on whether one is speaking to the addressee in person or communicating by written message (Clark & Brennan, 1991). Furthermore, the relationship with a dialogue partner affects our communication. The wording varies according to the audience, whether it is a friend, parent, or a work colleague (Fujishin, 2019). Communication involves not only words but also intonation, speaking rate, volume, pauses, and nonverbal expressions such as laughter, throat clearing, or an annoyed groan (DeVito, 2000). Anyone who has had experience transcribing interviews can imagine how important this non-vocal information is for correctly assessing the context of a statement. When trying to write a dialog with an IVA, these resources can be used to make the interaction more interesting and playful.

In addition to such stylistic considerations, several guidelines (Bouzid & Ma, 2022; Garcia et al., 2013; Murad et al., 2018) can help when writing a dialogue with an IVA. For interaction with an IVA to be successful, the user must first be made aware of the scope of the interaction with the IVA. If the scope is clear, it helps the user understand what can be asked and answered, i.e., what ontology is likely to be applied to the interaction (Bouzid & Ma, 2022).

The design of a VUI should be kept as minimalistic as possible to avoid overwhelming the user (Murad et al., 2018). Menus, for example, should not contain more than five or six items so that the user can remember them (Bouzid & Ma, 2022), as there is no screen to display the options. Furthermore, when working with sounds in addition to speech, one

has to think very carefully about when and how long a sound should be heard. Too much input at once overloads the VUI and makes it difficult to interpret the meaning of each sound (Röber & Masuch, 2005).

Since, unlike a GUI, oral information transmitted cannot be viewed once it's shared. Therefore, it should always be ensured that the information can be repeated (Garcia et al., 2013). Suppose a piece of information is repeated, either due to a query by the user or a similar query as before. In that case, the conversation becomes more engaging if the wording is not always the same. Imagine the following dialogue from a learning scenario similar to the one developed for the field study in chapter 5:

```
Alexa: "What is the translation of 'heute'?"  
User:  "The correct answer is today."  
Alexa: "That's right, here comes the next question [...]"
```

If the wording is not varied when the learning content is queried, the listener quickly becomes bored and frustrated. Thus, it is recommended to provide various sets of answers that can be modified according to the requirements. Even small rearrangements of words can lighten up the conversation.

Feedback is crucial for successful VUI design (Murad et al., 2018). In the absence of visual information, the user must be guided by audio feedback. Factors such as the time it takes for a VUI to provide feedback to the user play an essential role (J. Wang et al., 2023). Positive or negative feedback can be used to convey whether the system can process the input. However, in the case of an error, care must be taken to make clear to the user exactly where the problem lies (Lee et al., 2019). Does, for example, the answer not fit the context, or was there a problem with language processing? Feedback can be designed in very different ways and greatly impacts how an IVA is evaluated by users (Lee et al., 2019). Although it is known that feedback is essential to the design of a VUI, there is still room for improvement in its implementation (Murad et al., 2021; Myers et al., 2018). One possible approach to overcome this challenge could be gamification.

2.3. (Video-)gaming Without Graphics

To establish a foundation for connecting gamification and IVAs, it is crucial to consider existing concepts implementing playful approaches without a screen. This chapter presents various examples that demonstrate existing possibilities of implementation.

Examples of video games are familiar to most, with elaborate 3D graphics, now even playable in virtual reality environments, or simpler games with 2D graphics, such as Candy Crush that can be played on a smartphone. These video games work with various visual elements that illustrate a virtual world and a narrative the game is dealing with. At the same time, players are shown their progress, an overview of what other tasks and goals can be pursued, and how the other players are doing. But there are also several examples of (video) games that work without elaborate graphics.

One of the first types of computer games was text adventures. These appeared in the 1970s and 1980s. In this type of computer game, the player's environment and actions are conveyed through a text-based user interface rather than graphics and sound. The player also interacts with the virtual world by entering text commands. One of the most popular and widely played text adventures is the game "Zork." Complex puzzles, a deep story, and interactive elements characterize it. Like simpler rule-based CUIs, the game provides a procedural definition of syntax and grammar so that player input can be recognized and responded to appropriately (Fencott, 2012).

An example of audio-based games played without a screen is radio games. These games are usually game shows where listeners can call in to answer quiz questions (Stachyra, 2012). The questions in such a quiz can be knowledge questions or guessing sounds or songs. The mechanics of these radio games also vary, from simple games where the listener calls in to answer a question and wins or loses to round-based games where the listener has to compete against several opponents. Often, there are prizes to be won.

Audio games are another example of computer games without visual elements. Most audio games are developed for people with visual impairments (Urbanek & Güldenpfennig, 2019). The interface of audio games is implemented purely auditory, making it challenging to convey all the information relevant to the game (Friberg & Gärdenfors, 2004). These games use sounds, noises, music, and voices to communicate information to the player. Players usually use headphones to play these games, and input is provided through a keyboard or via touch-based gestures on a smartphone or tablet (Urbanek & Güldenpfennig, 2019). Voice is often used in these games to provide orientation, e.g., by a narrator describing the current scene or summarizing the events (Röber & Masuch, 2005).

With the fast development of IVAs and the possibility for developers to market their own skills via the platforms of IVA providers, multiple games have been released. Games for IVAs are almost always voice-controlled, though there are exceptions, such as Amazon's Echo Show, which has a screen that can be used in games. How this type of game can be described and what elements distinguish it will be discussed in chapter 6.

The various examples listed here show the versatility of playing games even without elaborate video graphics. Looking at existing game designs when developing gamification concepts is also helpful. This, along with the various approaches presented, provides hope that implementing audio-gamification is a beneficial approach to assist IVA applications in providing motivational design strategies.

CHAPTER

THREE

Impact of Individual Game Design Elements

This chapter is based on the results of a study published in the following peer-reviewed journal article:

Mazarakis, A., & Bräuer, P. (2023). Gamification is Working, but Which One Exactly? Results from an Experiment with Four Game Design Elements. *International Journal of Human-Computer Interaction*, 39(3), 612–627. <https://doi.org/10.1080/10447318.2022.2041909>

Impact Factor 2022: 4.7

The study was conducted in the frame of a seminar. The research question, hypotheses, and research design were developed together with Dr. Athanasios Mazarakis. Data collection was conducted by the author. Data analysis and manuscript preparation was done equally by the author and Dr. Athanasios Mazarakis.

3.1. Preface to Study 1

Gamification has already been proven to motivate users in many studies and practice (Seaborn & Fels, 2015; Warmelink et al., 2018). However, the implementation of gamification is a complex task that should not be underestimated (Morschheuser et al., 2017). Designers face difficult decisions, especially when trying to incentivize usage of the IVA while not overloading the application with a gamification concept (Doney, 2019). One way to address this challenge is to investigate whether simple concepts or even individual game design elements are sufficient to unleash the motivating effect of gamification. In this chapter, this research question will be addressed.

In the context of the thesis, this publication represents the first step in examining the effect of individual game design elements in detail. The findings from the study create an understanding that it is not always necessary to have a complex game design consisting of a multitude of game design elements to motivate users. Among other aspects, this is crucial for applying gamification in the context of IVAs. Due to the limitation of audio output in the implementation of applications for IVAs, any additional information that needs to be conveyed is a hurdle. To not overwhelm the user, the information must be kept short (Bouzid & Ma, 2022). This would be counteracted by incorporating complex gamification concepts consisting of various game design elements. Therefore, considering gamification on the level of individual game design elements is an essential prerequisite for later investigating gamification in the context of IVAs.

3.2. Study 1 (publication)

Gamification is Working, but Which One Exactly? Results from an Experiment with Four Game Design Elements

Athanasios Mazarakis  and Paula Bräuer 

Department of Computer Science, Kiel University, Kiel, Germany

ABSTRACT

Current gamification research usually examines several game design elements at the same time, which makes it difficult to distinguish how and to what extent individual game design elements increase motivation. We address this research question by individually examining four game design elements (progress bar, narrative, feedback, and badges) in an online experiment. In addition, combinations of game design elements were tested to gain insight about additive effects on motivation. The study included 505 subjects who answered a maximum of 190 different multiple-choice questions. The subjects were told to answer questions only as long as they enjoyed answering them. The results provide statistically significant motivational gains for all individual game design elements. Interestingly, not all game design elements benefit from a combination in the same way. The results of our study indicate that an increase in motivation through gamification is already possible if only an individual game design element is added.

1. Introduction

Gamification is a concept that is used to increase motivation to perform certain tasks (Seaborn & Fels, 2015) and is defined as the use of game design elements in a non-game context (Deterding et al., 2011). A non-game context may include, for example, systems to promote fitness, well-being and healthcare (Johnson et al., 2016; Miller et al., 2016; Sardi et al., 2017), school education (Conrad et al., 2010; Dicheva et al., 2015; Domínguez et al., 2013; Hanus & Fox, 2015; Majuri et al., 2018), working environments (Huschens et al., 2019; Sailer, Hense, Mandl, et al., 2017; Warmelink et al., 2018) or are designed for special groups like scientists (Feger et al., 2018, 2019; Mazarakis & Bräuer, 2020a, 2020b) or for inclusion of individuals with special needs (Patzer et al., 2018). The concept of gamification is recognized in the field of human-computer interaction (Rapp et al., 2019).

Gamification is inspired by video games, which possess the ability to have an intrinsic motivating effect by addressing the basic psychological needs for autonomy, competence and relatedness (Ryan et al., 2006). It is therefore reasonable to transfer this motivating effect of video games to other non-playing contexts via gamification (Seaborn & Fels, 2015).

The definition of Deterding et al. (2011) is not the only existing definition about gamification. Another definition that applies to the context of this article defines gamification as the use of “[...] game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning, and solve problems.” (Kapp, 2012, p. 54). Many

more exist, but the focus of each definition differs and is shifted toward the general idea of gamification (Muntean, 2011), business benefits (Werbach & Hunter, 2012), user experience and engagement (Domínguez et al., 2013) or a path to mastery (Kumar, 2013). Many of these definitions may be more appropriate in individual cases, but the definition of Deterding et al. (2011) seems to be most helpful with regard to making it possible to generalize. A further discussion of the issue of defining gamification can be found in Treiblmaier et al. (2018).

Gamification can also be considered as a technique to create or promote flow condition (Hamari & Koivisto, 2014). In contrast to flow theory, where this condition requires full focus to get into flow (Csikszentmihalyi, 2009), gamification, on the other hand, integrates game design elements in a way that they are perceived both consciously and unconsciously.

Countless apps and online platforms are now using game design elements to make their applications more motivating and retain customers (Kalafatoğlu, 2020), like for example Google Maps with their local guide system (Deal, 2018). Various studies have already shown that gamification has in general a motivating effect (Dichev & Dicheva, 2017; Hamari et al., 2014). Usually, in order to test the effect of gamification in different situations, most studies use a combination of different game design elements simultaneously (Domínguez et al., 2013; Sailer et al., 2013). Furthermore, the results achieved by using game design elements can differ in great extent (Koivisto & Hamari, 2019). Still, not every

gamification project is successful. Some projects have failed because, e.g., the gamified system was too intrusive, was not aligned to already existing systems or in general was below the initial expectations (Liu et al., 2017). From this point of view, gamification is no guarantee of success.

A possible explanation for these differences could be attributed to an interaction of the different game design elements. This means that individual game design elements could reinforce or neutralize each other in their effect. In order to find out whether and how different game design elements might influence each other, the first step is to look more closely at the effect of the individual game design elements (Dichev & Dicheva, 2017). But how exactly different game design elements work and whether there may be differences between the individual elements in terms of supporting motivation, has so far only been investigated by a few studies (Bräuer & Mazarakis, 2019; Christy & Fox, 2014; Groening & Binnewies, 2019; Hamari, 2017; Mekler et al., 2017). This is an open research question for many years (Mekler et al., 2013) and it is still not answered sufficiently (Koivisto & Hamari, 2019; Mazarakis, 2021).

This experimental field study addresses this question and examines individual game design elements with regard to their motivating potential. In our study we examine four game design elements (badges, feedback, progress bar and narrative) separately. We conducted an online quiz which was gamified with the game design elements mentioned above, testing them individually and in combination. The term quiz is complex and is defined by Raikar (2021) as “a contest in which participants test what they know by answering questions on one or more topics,” which means that a quiz is not necessarily a game. However, at the same time, “the term quiz is a capacious one” because, for example, “it can refer to a single game consisting of just a few questions, or it can refer to a large-scale event involving dozens or hundreds of people.” (Raikar, 2021). A non-game example can be considered when students are taking exams in the field of medicine at a university (Dengri et al., 2021) or when completing an online training course’s quiz with enterprise gamification (Stanculescu et al., 2016). Furthermore, a quiz can be considered as a game element (Saxegaard & Divitini, 2019), a learning element (Berger et al., 2019) or as an affordance (Koivisto & Hamari, 2019) in gamification. For example, an online quiz to examine the effects of gamification on learning was also used in a study by Sanchez et al. (2020), showing positive effects for short-term assignments. In general, gamification can be considered as conceptionally close to game design, however at the same time a gamified application does not have to be fun (Landers et al., 2018, p. 317).

The results of the experiment provide new and valuable insight into the individual and combined effects of gamification on motivation. This is contrary to the general approach of obtaining self-reported measures through storyboards (Hallifax et al., 2019), questionnaires or surveys (Koivisto & Hamari, 2019), which are usually easier to conduct but also provide less insight, because these findings are less comparable in contrast to well-executed inferential studies like

experiments (Hamari et al., 2014). To advance gamification research we need controlled experiments in order to gain knowledge on the actual effects of gamification.

After the introduction section we present related work regarding motivating individuals, fundamental challenges in gamification research and a general description of the game design elements we used in the study. The following methods section includes the hypotheses, a detailed explanation of the design of the gamification and the experiment itself. After describing the sample, descriptive and inferential results are given. We conclude this article including sections about discussing the findings, limitations, and outlook for future research. The results of this study are based on a poster presented at the conference “Technology, Mind, and Society (TechMindSociety ’18)” (Mazarakis & Bräuer, 2018). This article is a completely enhanced and extended version of the abstract published.

2. Related work

2.1. Motivating individuals

As Ryan and Deci (2000) point out, intrinsically motivated individuals perform an activity because the activity itself is interesting, rather than because of a separate consequence, reward or pressure. However, not every individual is identically intrinsically motivated by an activity, but there are individual and contextual differences in motivation potential. On the other hand, extrinsically motivated individuals perform an activity because of the associated remuneration. Thus, extrinsic motivation is the opposite of intrinsic motivation. Hereby the intrinsic motivation is attributed usually with a higher potential. In addition, extrinsic motivation is necessary, because many activities possess no intrinsic motivation potential, like for example a boring task (Ryan & Deci, 2000).

Reaching and sustaining intrinsic motivation is not an easy task, because many risks threaten it. For example, competition can jeopardize intrinsic motivation (Vansteenkiste & Deci, 2003), whereby personality traits can have an influential character (Epstein & Harackiewicz, 1992; Tauer & Harackiewicz, 1999; Vallerand et al., 1986). In an experiment it was shown that high performance motivated subjects prefer a competitive environment, whereas low performance motivated subjects prefer to avoid competition (Tauer & Harackiewicz, 1999). In addition, the actual performance of the subjects is also influenced. Competition can therefore be a social phenomenon that can both increase and decrease intrinsic motivation (Epstein & Harackiewicz, 1992), so that we need to consider trait and state of individuals. Thus, it becomes clear that the motivation of individuals is an overly complex matter whose investigation faces many challenges.

Gamification research tries to avoid a clear positioning of whether gamification is intrinsic or extrinsic motivation by using the term “affordances” (Jia et al., 2016). In line with the definition of motivation of Ryan et al. (2006), we consider gamification as externally endorsed and therefore extrinsic. Mekler et al. (2013) also assume that while game design elements can create an impact on intrinsic

motivation, this effect is extrinsically generated. However, it cannot be ruled out that in some cases an extrinsic motivation may be transformed into an intrinsic motivation.

2.2. Fundamental challenges in gamification research

Many experimental studies in the field of gamification have no individual consideration of the used game design elements, but usually all game design elements are applied simultaneously, leaving unanswered the individual effect of game design elements (Mekler et al., 2013). Although it has been shown in many studies that gamification can motivate in various ways (Hamari et al., 2014; Sailer et al., 2013; Seaborn & Fels, 2015), we argue the same way as Mekler et al. (2013) that possible effects of individual elements remain often undetected. This is necessary to provide support for good design choices, which is also acknowledged as a research gap (Koivisto & Hamari, 2019). Still, there are at present only very few experimental studies (Christy & Fox, 2014; Huschens et al., 2019; Landers et al., 2017; Sailer, Hense, Mayr, et al., 2017) that consider the effects of individual game design elements.

Additionally the triad of points, badges and leaderboards (Werbach & Hunter, 2012) are still the most frequently examined game design elements in gamification research (Koivisto & Hamari, 2019). Mekler et al. (2013) could show that points, levels and leaderboards, applied individually in an online image classification process are viable means to influence user behavior. It is therefore not necessary to apply all game design elements together. It is also assumed that individual game elements fulfill different functions or address different basic needs. However, the motivating effect could be different in interaction with other elements (Sailer et al., 2013). Therefore it is important to be aware of the relationship between the different elements.

Another study, which addresses individual game design elements and the combination of different elements, was carried out by Schöbel et al. (2016). By conducting a survey, they examined which game design elements and combinations are preferred in learning environments. Most popular was the game design element level followed by points and goals. With regard to the number of combined elements, most subjects indicated that they preferred between 3 and 4 game design elements (Schöbel et al., 2016). However, these results should be viewed with caution since they are not based on an experimental study but on a survey.

To get experimental and empirical insight into the individual effects of game design elements, we need to limit our study to a certain number of elements. So, we focus on the game design elements badges, feedback, progress bar and narrative. On the one hand we approach already heavily researched game design elements like badges and feedback (Hamari, 2017; Mekler et al., 2013; Seaborn & Fels, 2015; Zichermann & Cunningham, 2011). On the other hand we focus additionally on game design elements that are rarely considered in gamification research like progress bar and narrative (Dicheva et al., 2015; Garcia-Marquez & Bauer, 2021; Koivisto & Hamari, 2019; Sümer & Aydın, 2018). Also

we intentionally did not research just the three game design elements points, badges and leaderboards, also known as the PBL triad, as it is already known that they can harm in multiple ways motivation and performance (Mekler et al., 2013, 2017), in particular leaderboards as prominent example (Bräuer & Mazarakis, 2019; Werbach & Hunter, 2012). In addition, Kapp (2014a) states that the “most effective gamification efforts include more than points and badges — they contain elements of story, challenge and continual feedback [...]” (Kapp, 2014a, p. 52).

The analysis of individual and joint game design elements is a current trend in gamification science, because “possible effects of individual elements often remain undetected” (Mazarakis, 2021, p. 283). In addition, most gamification studies either compare absent vs. present gamification and do not control for the number of game design elements (Groening & Binnewies, 2021, p. 3). Furthermore, in their seminal article Hamari et al. (2014) show that on average 2.30 game design elements have been used in experiments with a standard deviation of 1.20. We use four game design elements in our study, which is more than one standard deviation above the average number of game design elements typically used in experiments.

The four game design elements examined in this study differ greatly in terms of the complexity of their design. Badges and in particular narratives usually require a lot of effort, because it is important to design them in a way that they appeal to the user. They need a careful design and need to consider usually mainstream taste. On the opposite to that, feedback and progress bars summarize information, which is usually easier to implement. We provide a general overview of the four game design elements (badges, feedback, progress bar and narrative) and will later detail in the procedure section how we realized these elements in our study.

2.2.1. Badges

Badges are virtual artefacts that are visually represented. Taken from the game design element “achievements,” they consist of three elements: signifier, completion logic and rewards (Hamari, 2017). They are awarded to the user for completing tasks (Antin & Churchill, 2011). Together with the game design elements points and levels, they are among the most frequently used game design elements in gamification research (Mekler et al., 2013; Werbach & Hunter, 2012). Badges can have different functions depending on how they are designed. So they can be used to create a comparison with others or to challenge oneself (Gibson et al., 2015). With the help of badges it is possible to specify certain target values for the user and to incentive the fulfilment of these (Sailer et al., 2013). Finally, a recent study asked subjects how they perceive badges. The results range from the perception of badges as rewards to social signaling and a function as goal setting to informing and encouraging subjects (van Roy et al., 2019). These results are partly in line with previous results (Antin & Churchill, 2011). The possibility to share badges can also motivate subjects additionally (Sheffler et al., 2020).

Badges can be implemented in different ways. To set a goal, the user is directly informed about which badges they can receive and how to act to get them. Further possibilities of application are badges where it is not known how they are awarded. The users only know that there exist badges, but they do not know what to do to get them. If a badge appears unexpectedly, this can cause surprise as well as joy for the user (Zichermann & Cunningham, 2011). Both approaches (known vs. unknown badges) are possible and reasonable to achieve a successful gamification of an application.

Also the game design element badges was examined in a study by Hamari (2017) individually. In a two-year study, Hamari (2017) was able to prove that users of a gamified version of a sharing platform were significantly more active than users without gamification. Hakulinen et al. (2015) could prove equally positive results on motivation and performance by the introduction of badges (Hakulinen et al., 2015). The authors looked at the effect of badges in an e-learning course and compared the behavior of a gamified group with a control group. Another study by Kyewski and Krämer (2018) also addresses the game design element badge. The authors examined the effect of badges in an online course. Unlike the other two studies, no effect on intrinsic motivation was generated by the badges (Kyewski & Krämer, 2018).

2.2.2. Feedback

Feedback is one of the most important game design elements for gamification (Kapp et al., 2014; Zichermann & Cunningham, 2011). In addition, with several different types it can be very valuable for gamification (Kapp, 2014b). A general definition states that feedback “[...] is information presented that allows comparison between an actual outcome and a desired outcome” (Mory, 2004, p. 746). Feedback is intended to provide the users with information about their performance or the status of their actions, which makes it possible to change behavior (Kapp, 2012), even just with a simple right/wrong feedback (Mazarakis, 2015), also known as confirmational feedback (Kapp, 2014b), to increase participation. Geelan et al. (2015) used feedback as one of two game design elements in an online learning tool. Based on a survey, a positive influence on the motivation and commitment of the participants during learning could be determined.

Additionally, the motivational construct of feedback plays an important role in the research area of intrinsic and extrinsic motivation. In a comprehensive study, it was shown that feedback alone can increase intrinsic motivation as well as, in combination with perceived autonomy and competence and eventually lead to an increase in performance (Ryan & Deci, 2000). Feedback is also crucial for flow theory. Only through feedback can individuals become aware of whether they are doing the activity in question as intended or not. Immediate and constant feedback can then help to maintain the flow state (Kapp, 2012). We also note that feedback can have different definitions and that it is an overly complex concept (Ramaprasad, 1983). In addition,

feedback can be distinguished between intrinsic and extrinsic feedback (Patchan & Puranik, 2016). Very generally speaking, Patchan and Puranik consider intrinsic feedback as feedback that occurs naturally. On the contrary, extrinsic feedback is feedback that is provided by an external source. Extrinsic feedback is also usually regarded just as “feedback” (Patchan & Puranik, 2016, p. 130). However, feedback can be acknowledged as one key element of gamification (Mazarakis, 2015; McGonigal, 2011).

It is also worth noting that feedback can be expected as a complimentary game design element for a quiz. To determine whether this is the case, anonymized responses collected *via* a form at the end of the experiment are evaluated.

2.2.3. Progress bar

Most video games (and probably also life itself) are about moving on in some way, developing further or achieving certain goals. One of the simplest elements that can be used to visualize this progress is a progress bar (Siemens et al., 2015). Progress bars are also used in numerous online portals and shops to motivate users to complete information on their profile pages or to guide them through the steps of a purchase process. So a subject gets a clear information about, e.g., how much percent of an activity or task has been accomplished and how much approximately is left (Myers, 1985). Unfortunately, this game design element is often neglected, despite its rather easy implementation to a system. So, missing research about progress bars as a game design makes a reliable assessment of its potential rather difficult. Both Dicheva et al. (2015) and Sümer and Aydın (2018) argue as well that progress bars are a niche in gamification research compared to other game design elements.

Progress bars can be used to represent objectives in a comprehensible way and also provide subjects a graphical information about the (partial) success of the objective (Sailer et al., 2013). Current studies investigating progress bars (Geelan et al., 2015) do not yet assess the individual effect of this game design element on motivation. Only recently there was an exception to this practice, where a study with 185 subjects in the context of the theory of gamified learning, mostly failed to find statistically significant results for progress bar and a combination of progress bar and badges (Garcia-Marquez & Bauer, 2021).

2.2.4. Narrative

Narrative as a game design element (also known as story or storytelling) is also essential for gamification, in particular when it comes to learning or giving instructions, as it gives meaning to things (Kapp, 2012). A narrative is usually a story that guides a user by providing textual or spoken information. It can be used to understand goals or give meaning to an activity or task.

In addition, a narrative can create or enhance a sense of social or emotional experience (Nah et al., 2013). Narratives can be used to add meaning to other game design elements and encourage users to become more involved (Grobely et al., 2018). In non-game contexts, narratives can play an

important role because they can change the meaning of activities in the real world (Sailer, Hense, Mayr, et al., 2017). In physical education, for example, climbing on equipment and skillful jumping from one place to another can be much more motivating when “the ground is made of lava.” When designing a narrative for a non-game context, care should be taken to create a reference to the real situation and to align the competence and meaning of the story with the context (Kapp, 2012).

3. Method

For this study, a field experiment was conducted with a between-subjects design. Each subject was at the beginning of the experiment randomly assigned to one condition, in which the subject remained for the entire experiment. Subsequently, voluntary information on age and gender was requested before the start of questions. The experiment comprised a total of seven experimental conditions. Besides the control group (CG) which did not consist of any game design element, six experimental groups with individual game design elements and combinations were created. Three of these experimental conditions were each equipped with an individual game design element: feedback (FB), progress bar (PB), and narrative (NR). We did not examine the badge game design element individually because there is a large number of research supporting the positive effect of badges for gamification (Antin & Churchill, 2011; Bräuer & Mazarakis, 2019; Gibson et al., 2015; Hakulinen et al., 2015; Hamari, 2017; Kyewski & Krämer, 2018; Mazarakis & Bräuer, 2018, 2020a).

A combination of the feedback element and another game design element was chosen for each of three further groups: progress bar and feedback (PB + FB), narrative and feedback (NR + FB) and badges and feedback (BA + FB). This gives us the opportunity to test both individual game design elements and combinations. The assumption when designing the combinations was that the feedback condition would have the weakest effect of the four elements, as this condition is most likely not to be associated as a game element, although it is indeed classified as such in the scientific literature (Geelan et al., 2015; Kapp, 2012, 2014a, 2014b; Mazarakis, 2015; Zichermann & Cunningham, 2011). An increase in motivation through the combination with one of the other game design elements, if any, would be most likely to be expected here. This also strengthens the validity of the individual elements because now a comparison is possible. Given that the individual groups and combinations have been defined, the corresponding hypotheses are set out in the following section.

3.1. Hypotheses

The research question of our experiment is to find out whether the number of questions answered in a quiz can be increased by using, *ceteris paribus*, different game design elements. Based on stated related work, the following six hypotheses are specified to address this research question.

The hypotheses H1 - H3 assume that a motivating effect can already be achieved by applying only one individual game design element. This assumption so far is not well enough studied in research (Mekler et al., 2013).

- Hypothesis 1 (H1): The subjects in the feedback condition answer more questions than the control group (FB > CG)
- Hypothesis 2 (H2): The subjects in the progress bar condition answer more questions than the control group (PB > CG)
- Hypothesis 3 (H3): The subjects in the narrative condition answer more questions than the control group (NR > CG)

Of further interest is whether additive effects may arise from the combination of several game design elements, which is a common practice in gamification research but often difficult to interpret (Sailer et al., 2013). The hypotheses H4 - H6 will shed light whether and to what extent a combination of several game design elements can efficiently support motivation. Usually the effect of feedback, in particular if it is just a right/wrong feedback, is taken for granted. With our study we can estimate the additional quantitative effect, if there is any, in comparison to the feedback condition alone.

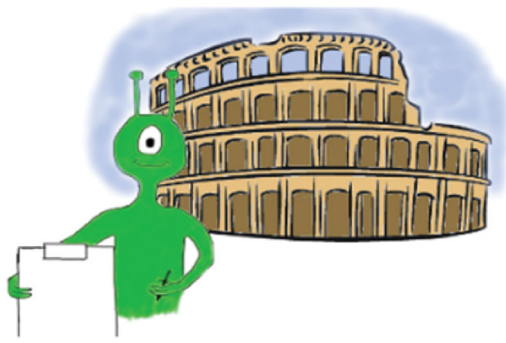
- Hypothesis 4 (H4): Subjects with the combination of progress bar and feedback will answer more questions than just subjects in the feedback condition (PB + FB > FB)
- Hypothesis 5 (H5): Subjects with the combination of narrative and feedback will answer more questions than just subjects in the feedback condition (NR + FB > FB)
- Hypothesis 6 (H6): Subjects with the combination of badge and feedback will answer more questions than just subjects in the feedback condition (BA + FB > FB)

It must be noted that the hypotheses H4, H5, and H6 are relevant to assess the additional gain by combining two game design elements. Previous statistical significance *vs.* the control group is assessed and necessary to check these three (H4, H5, and H6) hypotheses.

3.2. Procedure

An online quiz was developed to experimentally test the hypotheses presented in the previous section. Basically, the quiz is about answering questions about continents, countries, and space. Some studies already used gamified quizzes or tests to investigate the effect of gamification (Bevins & Howard, 2018; Cheong et al., 2013). An advantage of our study design as a between-group experiment is that the motivation of the subjects can be objectively measured and compared and is not based on a subjective self-assessment via a survey.

The quiz consists in total of 190 questions. Each question is provided with four possible answers but only one is correct. The questions of the quiz are divided into nine



During the eruption of which volcano was the Italian city of Pompeii destroyed?

- Etna
- Monte Vulture
- Vesuv
- Marsili

Next question

Quit

Figure 1. Sample question from the narrative condition.

thematic blocks. Seven blocks focus on ten general questions each about the seven continents (in total 70 questions). Also we asked five questions each about four different countries of the continent (Europe: Germany, Greece, Italy and United Kingdom; Asia: China, India, Japan and Russia; Africa: Egypt, Nigeria, South Africa and Madagascar; North America: Canada, USA, Mexico and Caribbean; South America: Brazil, Ecuador, Argentina and Peru) excluding the two continents Australia and Antarctica (in total 100 questions). The final two sections of the quiz each contain ten questions about the moon and space (in total 20 questions). All questions are displayed in each condition in the same order to ensure comparability between the experimental conditions. An example of a question is shown in Figure 1.

The subjects are not informed of the total number of questions in the quiz in any of the conditions. The basic idea of the quiz is to let the subjects answer questions as long as it is fun for them, in order to measure the effect of the various game design elements on the basis of the number of questions answered.

If a subject wants to finish the quiz, then this is possible by pressing the “Quit”-button. The quiz also is finished by answering all 190 questions. It is not possible to pause the quiz and to resume later. After completing the quiz, the subjects were told how many questions they had answered in total and how many of their answers were correct. This was not communicated to the subjects in advance in any experimental condition, so as not to motivate them in addition to the game design elements for the conduction of the quiz. In addition, the subjects had the opportunity to provide anonymous comments via a form at the end of the experiment.

What is the shape of the Earth's solar orbit?

Circle

Ellipse

Cone

Spiral

Next question

Quit

Figure 2. Sample question from the feedback condition.

3.3. Design of conditions and gamification

In this section we explain in detail the different conditions and the design of the gamification. Also, we provide examples for the experimental conditions.

3.3.1. Control group

The control group did not include any of the following game design elements (feedback, badges, narrative and progress bar). For clarification, there was no feedback if a question was answered correctly or incorrectly, as this is considered a game design element in this experiment (Kapp, 2012, 2014b; Mazarakis, 2015; Zichermann & Cunningham, 2011). Using an unmixed control group, it is possible to calculate the actual effect of each game design element. In addition, the intrinsic value of the quiz is made visible.

3.3.2. Feedback















For the implementation of the feedback element a “right-wrong” feedback is chosen. This can be considered as extrinsic feedback because our “right-wrong” feedback is provided by an external source and does not occur naturally (Patchan & Puranik, 2016). After a subject confirms their chosen answer to a question by clicking the “Next question” button, the answer is immediately colored green (correct answer) or red (wrong answer). In addition, if the answer is incorrect, the correct solution is highlighted in green color. Figure 2 gives a visual example of the feedback condition. This is in line with confirmational feedback, considered to be a game design element (Kapp, 2014b; Kapp et al., 2014; Mazarakis, 2015).

3.3.3. Badges

For the experimental condition badges 14 different badges were designed. They can be seen in Table 1.

Eight of the 14 badges were only awarded if a specific question has been answered correctly. One of the badges was implemented as a level badge (bronze/silver/gold), as it is often found in other studies (Hamari, 2013; Kyewski & Krämer, 2018) and games. The remaining badges were unlocked when a particular question was answered correctly or for several correct answers in a row. At first, the badges are only visible as 14 gray circles displayed in a box above the current question. Before the subjects unlock one of the badges, it is not clear to them that they can obtain a badge.

Table 1. List of all badges with the condition for unlocking them and number of subjects who received them.

| Badge | Conditions for unlocking the badge | Number of subjects who received this badge |
|---|---|--|
|  | Answer any question correctly within 5 seconds | 40 (70.18 %) |
|  | All questions about a country within 50 seconds correctly answered | 44 (77.19 %) |
|  | All questions about a continent answered correctly within 150 seconds | 36 (63.16 %) |
|  | 3 questions answered correctly one after the other | 51 (89.47 %) |
|  | 5 questions answered correctly one after the other | 43 (75.44 %) |
|  | 10 questions answered correctly one after the other | 11 (19.30 %) |
|  | Question how often Germany has won the football world championship correctly answered | 35 (61.40 %) |
|  | Question who invented the light bulb answered correctly | 24 (42.11 %) |
|  | Question correctly answered that Antarctica has no official flag | 17 (29.83 %) |
|  | Answered question about Machu Picchu correctly | 22 (38.60 %) |
|  | Question about the Terracotta Army correctly answered | 15 (26.32 %) |
|  | Question about Africa's export products correctly answered | 12 (21.05 %) |
|  | Question about the world wonder Zeus statue answered correctly | 13 (22.81 %) |
|  | Question about the number of time zones in Australia correctly answered | 17 (29.83 %) |

In addition, it is not clear what conditions the subject has to meet to get a badge. This approach was chosen to incentivize only the behavior of answering questions and not the effort to achieve a goal. If we would incentivize the targeted collection of badges through a planned behavior by the subjects, we would be conforming with the ideas of the goal setting theory (Locke, 2001; Locke & Latham, 2002). The goals set by the badges could then in turn lead to subjects working precisely up to a certain question to achieve a certain badge. On the one hand, this behavior would counteract the experimental design pursued in this study, in which the motivation of the subjects is measured based on the questions answered. On the other hand goals are implicit components of other game design elements like achievements and quests (Werbach & Hunter, 2012), which could be perceived by individuals as another game design element which confound randomly with our approach, to research the effect of individual game design elements.

The design of the badges is aligned with the findings of Facey-Shaw et al. (2018). As suggested by the authors, various linear and non-linear badges that may be acquired progressively after achieving specific goals are incorporated in

the badge design of our study (Facey-Shaw et al., 2018, p. 538). This is implemented in our experiment by providing badges for each continent (linear) and badges without relation to a continent (non-linear). In addition, we consider for our badge design two aspects as relevant. First, we want to support that subjects can achieve mastery (Abramovich et al., 2013). This is the case for the first six badges in Table 1. Moreover, second, learning goals and reinforced learning shall be supported in an unspecific manner by the 14 badges (Botra et al., 2014; Davidson & Candy, 2016). So, in general, we designed the badges to reflect advancement as milestone badges and as outcome-based badges when answering some specific questions correctly (Facey-Shaw et al., 2018, p. 539).

Finally, we were careful about the visual representation of the badges, although visual appeal may not be significant in terms of success of the gamification, but it still can have an impact on interaction with the gamified system (Facey-Shaw et al., 2018, p. 540). After receiving a badge, one could find out why a badge was awarded by hovering the mouse pointer over the badge. As soon as the mouse hovers over one of the displayed icons, a tooltip field appears, which explains what the respective badge represents. In addition, a

pop-up window informs the subjects that they have just received a badge as soon as they unlock the badge.

3.3.4. Narrative

As Keusch and Zhang (2017) assume based on their literature review, when using a narrative for gamification, it should be chosen as appropriate to the topic as possible. Based on this assumption, a story with an alien was chosen to meet the topic “Earth and Space.” This alien was sent to Earth by a space commission to check whether planet Earth should give way to a space highway.¹ The subject is now asked to answer questions to convince the alien that the Earth is worth to be saved. Suitable pictures were created for all continents and countries thematized in the questions, which did support the story visually. When switching to the next continent or country, a new image and text were displayed with an introductory commentary by the alien. As an intermediate event a block of questions about the moon was added. The alien tells the subject that the moon is about to be blown up to make room for the construction engines. To prevent this, again questions must be answered. The intermediate event should make the story more engaging.

The subject receives no information about the outcome of the story. This is to avoid that subjects participate again in the quiz to change the outcome of the story. Though this would not harm the subject, the data gathered would not be suitable for analysis because the subjects would have some previous knowledge of the questions and maybe also remember the correct answers. This is undesirable for the purpose of our study, to find out the motivational potential of individual and combined game design elements. Kapp (2012) also notes that the integration and design of a narrative does not depend on the end of the story but on the journey through it. Figure 1 shows an example how the narrative looks like, including a picture of the alien.

3.3.5. Progress bar

For the progress bar, a round design in the form of a filling globe is chosen, closely related to the topic of the quiz. The image of the earth is uncovered in 20 steps. The subjects see at the top of the progress bar how many questions they still have to answer before the next piece of the globe is uncovered. The number of questions necessary for the next piece is not increased linearly but varies between 3 and 15 questions. In addition, the design of intervals of different size was also chosen to achieve higher motivation through infrequent progress. In contrast to regular progress, irregular progress has a higher motivation potential in gamification (Zichermann & Cunningham, 2011). Table 2 shows the numbered interval and the necessary number of questions to answer are displayed.

In addition, this procedure helped to hide the total number of questions available in the quiz and so not to provide an additional motivation element, which can be viewed again as goal setting (Locke, 2001; Locke & Latham, 2002). In total, the 190 questions were distributed over 20 pieces of the progress bar. Figure 3 shows a quiz question in the progress bar condition.

Table 2. Intervals and questions necessary to answer for progress in the progress bar condition.

| Interval | Questions to answers |
|----------|----------------------|
| 1 | 3 |
| 2 | 4 |
| 3 | 8 |
| 4 | 6 |
| 5 | 7 |
| 6 | 8 |
| 7 | 9 |
| 8 | 10 |
| 9 | 11 |
| 10 | 12 |
| 11 | 13 |
| 12 | 7 |
| 13 | 12 |
| 14 | 12 |
| 15 | 15 |
| 16 | 13 |
| 17 | 11 |
| 18 | 8 |
| 19 | 12 |
| 20 | 9 |



Figure 3. Sample question of the progress bar condition.

3.4. Subjects

The experiment was carried out for one month and subjects were recruited online via social networks, such as Facebook and Xing. No remuneration was paid for their participation, nor was any such remuneration advertised. In the end 531 participants took part in the study. Of these, 20 were removed due to duplicate IP addresses and session IDs. These subjects were excluded because it could be assumed that the subjects took part in the experiment a second time. However, the results of further participation cannot be evaluated because, firstly, there is the possibility that the subjects may be assigned randomly to a different experimental condition. This would lead to a confounding of the results

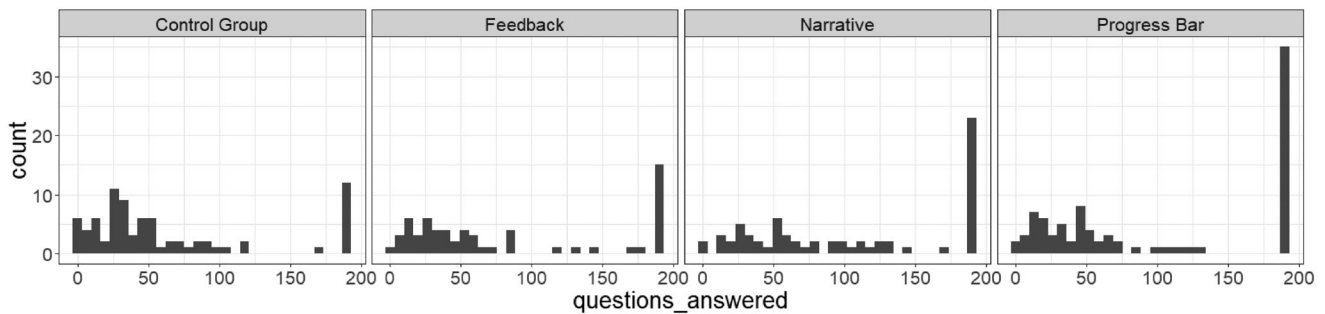


Figure 4. Frequency distribution of answered questions per condition/group for the individual game design elements.

because the respective subject could now recognize that there is more than one experimental condition and thus react differently. Secondly, it can be assumed that an effect on the subject's motivation can occur if the subject already knows the questions in the quiz. So, they could be motivated if they can answer more questions correctly or they could be demotivated if the same questions always appear in the same order. Because the direction of the effect of such events is not reliable estimable, the complete removal of these subjects is the only way to interpret the results objectively. Six further entries were excluded from the analysis due to contradictory information (like for example subjects that stated, that they are 99 years old), as well as additional conspicuous behavior, such as going too fast through the experiment without paying attention to the actual questions. The exculpated subjects were distributed over the seven experimental conditions as follows: 5 CG, 4 FB, 3 PB, 6 PB + FB, 3 NR, 2 NR + FB, and 3 BA + FB. The distribution of these 26 subjects is almost equally distributed without any statistically significant abnormality. This left a total of 505 subjects for subsequent statistical evaluation.

The allocation of subjects to the experimental conditions was randomized and each subject stayed in its experimental condition. It was not possible for the subject to change the group. Since all subjects reached the quiz via the same internet URL, the subjects were not aware that there was more than one condition available. At the beginning of the quiz, the subjects were informed about the anonymization of the data and about the fact that a "Quit"-button had to be pressed to send the data to the server. It was also pointed out that questions of the quiz should only be answered as long as it was fun for the individuals. This was also pointed out when advertising the study to ensure that each subject was aware that the quiz did not have to be fully answered. Subsequently, voluntary information on age and gender was requested before the start of questions. After finishing the quiz, the subjects were also given the option to provide an open comment about the study via a form.

4. Results

Our sample consists of 61% (308) women and 34% (172) men, 5% (25) did not provide a gender statement. The mean age for all subjects is 31.47 years (SD 12.98). The median is 27 years, and the age span starts from 14 to 78 years. The differences in age between women and

men were analyzed and are not statistically significant ($t_{(477)} = 0.32$, $p = 0.750$).

The aim of the study was to compare the motivation of the subjects between the different experimental conditions. The motivation of the subjects was measured by the number of questions answered. First the descriptive results and then the inferential statistical results are presented.

4.1. Descriptive results

The distribution of the subjects in each of the experimental conditions is shown in Figures 4, 5. We can observe from these distributions after how many questions in each condition the subjects quit the quiz. Obviously, in all conditions many subjects either quit the quiz early (after about 30 to 50 questions) or they answered all 190 questions, even in the control group.

Exceptionally striking are the two experimental conditions with the progress bar (PB, PB + FB) and the group with the badges (BA + FB). Exactly 50.0% of the subjects in the condition with the progress bar and feedback answered all questions, whereas this is the case for 39.3% in the condition with only the progress bar. In the badge and feedback condition it is also remarkable that 41.4% of the subjects answered all 190 questions. It can already be stated that the progress bar and the badges, always in combination with feedback, have motivated best the subjects to answer all questions.

The mean value of answered questions over all conditions is 98.79, with a standard deviation of 74.29. A total of 180 subjects (35.64%) answered all questions in the quiz. Table 3 shows the number of subjects, mean value of answered questions, associated standard deviation (SD), p -value and the number of subjects who answered all questions per condition. On average, the subjects answered 56.37% of the questions correctly. Because the questions were in general not so easy to answer and kind of challenging, it was examined whether the proportion of correctly answered questions might have had an influence on the point at which the quiz was finished by a subject. It can already be stated in advance that because there is no normal distribution of the data, a Spearman correlation was calculated to explore this aspect. The analysis reveals a low correlation without statistical significance for the number of questions answered and the proportion of questions answered correctly by an individual, $r_s = 0.08$, $p = 0.062$. Thus, it can be assumed that the

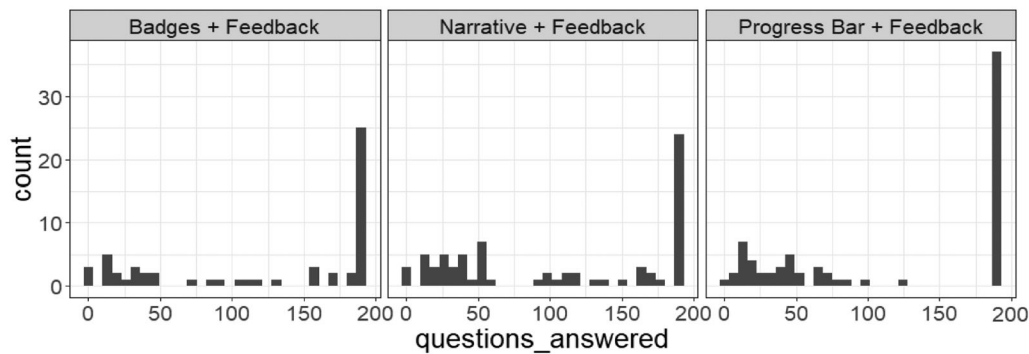


Figure 5. Frequency distribution of answered questions per condition/group for the combined game design elements.

Table 3. Number, mean value, standard deviation (SD), p -value, and the number of subjects with all questions answered per condition.

| Condition | N | Mean | SD | All questions answered | Proportion of questions answered correctly | p -Value |
|-----------|-----|--------|-------|------------------------|--|------------|
| CG | 80 | 63.35 | 61.58 | 12 (15.00%) | 56.95% | – |
| FB | 62 | 82.61 | 71.07 | 24 (38.71%) | 57.40% | 0.046* |
| PB | 89 | 101.21 | 76.04 | 35 (39.33%) | 56.86% | 0.000*** |
| PB + FB | 74 | 114.78 | 78.44 | 37 (50.00%) | 58.41% | 0.007** |
| NR | 69 | 105.59 | 69.16 | 23 (33.33%) | 58.65% | 0.000*** |
| NR + FB | 74 | 106.39 | 73.39 | 24 (32.43%) | 54.13% | 0.029* |
| BS + FB | 57 | 123.44 | 75.54 | 25 (43.86%) | 51.28% | 0.002** |
| Total | 505 | 98.79 | 74.29 | 180 (35.64%) | 56.37% | – |

* $p < .05$, ** $p < .01$, *** $p < .001$.

difficulty of the questions had no significant influence on the motivation of the subjects.

The mean value of unlocked badges was 6.67 (SD 3.68). Four subjects did not unlock any of the badges and four other subjects managed to get 12 out of 14 possible badges. All 14 badges were not obtained by any of the 57 subjects in the corresponding group. Table 1 also shows how often each badge was obtained.

It can already be seen from the descriptive results that the subjects in the control group answered fewer questions on average than in the other experimental conditions. For a more precise and reliable interpretation of the results, the inferential statistical analysis follows.

4.2. Inferential results

An analysis of variance is conducted for the statistical evaluation of the results to be able to interpret the mean value comparisons between the different groups. The test for the homogeneity of the variances (Levene test) for the number of answered questions is statistically significant with $p = 0.000$, whereby the Levene statistic is 7.37. All subsequent results are therefore based on unequal variances and corrected accordingly conservatively. This also leads to crooked values for the degrees of freedom.

The analysis of variance indicates a statistically significant difference between the individual experimental conditions, $F_{(6, 498)} = 5.68$, $p = 0.000$. Since the homogeneity of the variances is not given, the results must be corrected by the Welch test (Field, 2009). After the adjustment, the result of the analysis of variance is 6.51, $p = 0.000$. Due to the statistically significant difference in the number of answered questions between the individual experimental conditions,

all hypotheses can now be statistically examined and tested one-tailed.

The comparison of the mean values of the feedback group and the control group (H1) provides a statistically significant result, $t_{(121.01)} = 1.70$, $p = 0.046$, $\Delta = 0.31$. Consequently, H1 can be supported and it can be assumed that the game design element feedback motivates the subjects.

The comparison of the mean value of the progress bar group and the mean value of the control group (H2) also shows a statistically significant result, $t_{(165.26)} = 3.57$, $p = 0.000$, $\Delta = 0.62$. Thus, it can be assumed that the progress bar has a motivating effect because the subjects answered more questions in the progress bar condition, H2 is supported.

The subjects in the condition with the narrative also answered statistically significantly more questions than those of the control condition (H3), $t_{(137.47)} = 3.91$, $p = 0.000$, $\Delta = 0.69$. Consequently, H3 can also be supported and it can be assumed that narratives also have a motivating effect. In summary in this study there is support that each individual game design element has a positive effect on motivation to answer questions in a quiz.

Now we examine the results for the combinations of game design elements. It is already obvious from the descriptive results, that the differences between the control group and the experimental conditions with the combinations are significant. Therefore, the presentation of these significances will be omitted and the analysis of the hypotheses H4, H5, and H6 will be performed immediately. A statistically significant difference between the mean values was found for the combination of progress bar and feedback compared to the condition with feedback only (H4), $t_{(133.15)} = 2.51$, $p = 0.007$, $\Delta = 0.84$. Also, the comparison

between the conditions with narrative and feedback and feedback only (H5) also showed a statistically significant difference, $t_{(131.19)} = 1.92$, $p = 0.029$, $\Delta = 0.70$. The third tested combination of badge and feedback also shows a statistically significant difference when comparing the mean value with the condition with feedback only (H6), $t_{(114.58)} = 3.03$, $p = 0.002$, $r = 0.27$. All three conditions that used another game design element in addition to the feedback element provide significant results when comparing the mean values with those of the group that used only the feedback element. Consequently, H4, H5, and H6 can also be supported. Combining two game design elements helped to gain more answered questions than feedback alone.

5. Discussion

The motivation for this study was to assess how individually applied game design elements affect subjects' motivation. We could show that individual game design elements can be enough, to motivate individuals by finding support for the corresponding hypotheses H1, H2, and H3. This needs to be considered because the creation of two or more additional game design elements is usually associated with additional costs. But we may not neglect that we could show that a combination of two game design elements can support the motivation of subjects even further, as we can support our hypotheses H4, H5, and H6.

It is still also questionable, to be in favor of two or more combined game design elements at the same time, because in comparison the additional mean number of questions answered in our experiment, is not high enough to justify this effort. We can show <1% increase if we compare the narrative with the narrative and feedback condition. For the progress bar *vs.* the progress bar and feedback condition it is at least a 13% increase. Careful consideration is necessary whether an increase of the motivation justifies the additional effort.

For the implementation of the narrative element, a story was specifically designed to be suitable to the theme of the questions. As the results have shown, the game design element proved its motivating effect. However, it must be strongly assumed that the effectiveness of a narrative depends very much if and how the context of the study is considered adequately (Kapp, 2012). Though context is not a unique aspect to be considered for narrative (Finckenhagen, 2017), it is important to have a story where most stakeholders are pleased. The goals of the gamified system and stakeholder objectives need to be aligned and at the same time one must keep in mind the constraints of the organization or the experimental setting (Richards et al., 2014).

Based on our results in particular the use of the progress bar and badges can be recommended to increase motivation. The results on the motivational effect of badges go hand in hand with most results of previous studies (Hakulinen et al., 2015; Hamari, 2017; Koivisto & Hamari, 2019).

6. Limitations and outlook

As with any other experiment, our study has some limitations. First, context is very important in gamification research (Finckenhagen, 2017; Richards et al., 2014). Not only gamification in general, but also the effect of the game design elements, depends strongly on the context, systems or in general on the environment in which they are used (Finckenhagen, 2017; Hallifax et al., 2019; Koivisto & Hamari, 2019; Richards et al., 2014). For example, the context is essentially different if we want to gamify online shopping *vs.* learning. In the first case someone wants to increase the website activity, while in the second case the emphasis is on keeping the learning focus high (Liu et al., 2017). This makes it evident that the results of this study cannot be generalized if the context is not considered. Stakeholders or organizations can differ, which can also influence the effect of gamification. Along the same lines, many activities are not intrinsically interesting and require significant effort and persistence. Maintaining long-term user commitment through gamification is a recognized challenge. Whether the shown effects would last over time in other real-life activities is unclear. And the transfer of a game design element from one context into another context does not necessarily lead to the same motivational experiences. We still believe that our sample with 505 subjects is big enough, to draw certain conclusions from the results, but they can vary depending on the context.

We have also refrained from analyzing the game design element badge individually, as the existing research is noticeably clear in this respect (Antin & Churchill, 2011; Bräuer & Mazarakis, 2019; Gibson et al., 2015; Hakulinen et al., 2015; Hamari, 2017; Kyewski & Krämer, 2018; Mazarakis & Bräuer, 2018, 2020a). Nevertheless, it must be noted, although it seems rather unlikely, because no individual testing for badges has been applied, it is theoretically possible that badges could either have had no positive effect in our experiment or maybe a much greater effect, without combination with the feedback game design element. Also, the possibility to share badges via social media should also be of interest for future research, as this might additionally enhance the effect of gamification (Sheffler et al., 2020) through public display or comparison with others. Finally, testing known *vs.* unknown badges is a promising research question in order to figure out if a combination of badges and goal setting is more successful than badges alone.

Further limitations offer additional potential for future research, e.g., the occurrence of the ceiling effect. When designing the study, it was expected that a quiz with 190 questions would consist of enough questions, so that most of the subjects would quit before reaching the final question. Contrary to this assumption, however, in most groups at least one-third of the subjects answered all questions. This can lead to an impairment of the results, since it is not possible to find out for these subjects at which point the motivation decreased. In future studies, the number of questions in such a quiz should therefore be significantly increased. To counteract the ceiling effect, we suggest for future studies a minimum of 300 questions. Of course, the results are still

valid, because they underpin the high potential for motivation, they only limit the possible interpretation range.

A concurrent shortcoming and strength of the experiment is the clear focus of the individual conditions. It was expected that the control group would answer only a few questions of the quiz and that almost no one would reach the end of the 190 questions. The task in the control group was just to answer question after question, without getting any information if the answers are correct or wrong and without the application of any game design element. Still, it looks like that answering a quiz possesses a high intrinsic value for subjects. On average, more than 62 questions were answered and 15% of the subjects in the control group answered all 190 questions. So even without knowing if the questions are answered correctly or not, the subjects did not perceive the task as demotivating. This is a key finding that must be considered when designing future quiz studies.

In scientific literature, a quiz can be a game element, a learning element, or an affordance in gamification (Berger et al., 2019; Koivisto & Hamari, 2019; Saxegaard & Divitini, 2019). It is possible that quiz settings could positively influence the effect of game design elements if we consider a quiz as well as a game design element, as it is assumed that game design elements influence each other (Sailer, Hense, Mandl, et al., 2017). However, we can only assume that this is the case. Nevertheless, even if this would be the case, our findings are still valid by applying a between-subjects and randomized experimental design, including the *ceteris paribus* clause to exclude additional interactions and interfering factors. Furthermore, because we have a quiz setting in all conditions, we can safely statistically control for this variable.

It is possible that the subjects of our study expected some game design elements as complimentary in a quiz setting. To answer this question, we examined the anonymous comments we received from the subjects at the end of the experiment via a form. Out of 505 subjects, only two subjects would have liked feedback (one in the control group and one in the progress bar conditions), and another subject would have preferred a summary feedback per continent (combined progress bar and feedback condition). As a result, we have no indication that the subjects, in general, expected the existence of certain game design elements.

Another limitation of the study occurs in the design of the progress bar. A main objective of the study was to hide the maximum possible questions available in the study. This has been done to avoid the appearance of goal setting (Locke, 2001; Locke & Latham, 2002). We wanted to avoid subjects trying to answer all the questions in the quiz, so that they do not fulfill an (implicit) obligation. Our research design can be considered as an effort to minimize the effect of goal setting in gamification to observe the effects of our different experimental conditions. However, by using the progress bar, a rough estimate of the total number of questions is possible, which needs to be considered in the interpretation of the results. It is inherent to a progress bar that a quantitative or visual indication of progress is displayed

and that it can be concluded or estimated when the maximum number of questions can be reached.

In addition, the experimental implementation of the badges condition is not as perfect as anticipated. Possible criticisms are the same as the limitation concerning the progress bar. While we have aimed not to provide explicit goals and information on how to achieve the badges, we may have unintentionally provided an implicit goal and information. The badges are initially only visible as 14 gray circles placed in a box above the current question. Since it is not conveyed what the subject must accomplish to receive a badge, it is not clear to the subject until the moment of activation that a badge can be achieved for the current activity. However, displaying the 14 gray circles might suggest to some subjects that a complete collection of all badges is possible. This, in turn, could be perceived as an implicit achievable goal, which then would function like the game design element collections (Werbach & Hunter, 2012, p. 80), and could also be seen as the game design element challenge, as, for example, in Friedrich et al. (2020, p. 348). To determine if this affected our study, we examined the anonymous comments received from the subjects at the end of the experiment. Fortunately, the analysis shows that there are no comments on this issue. However, this is, of course, not definitive proof that there may not have been any issues. Still, there is no circumstantial evidence for this issue. To avoid this problem and to ensure that only the effect of the badges is measured, an implementation without gray circles would be preferable in future studies.

As in most gamification studies, the effect of the game design elements was only tested over a short period of time. So far, very few experimental studies with big samples, and not just survey studies (Warmelink et al., 2020; Xi & Hamari, 2019), have been carried out about the long-term effects of gamification, and in addition these provide mixed results (Hamari, 2017; Hanus & Fox, 2015). How the different individual game design elements perform over a longer period remains an open question that should be addressed in follow-up experimental studies.

Although it was not the actual intention, this study was able to show that by adding a second game design element, the motivation of the subjects could be additionally increased. However, only the combination of feedback and another game design element (badges, narrative or progress bar) was examined. In addition, it should be considered that a saturation effect may occur or that the motivating effect turns negative when the number of combined game design elements exceeds a certain point. Future research is needed to systematically analyze interaction effects between the individual game design elements and combinations.

Other game design elements may or may not provide same or different results, e.g., rankings or leaderboards. However, the state of research on other game design elements is not consistent. Huschens et al. (2019) investigated the effects of introducing rankings in a working environment (Huschens et al., 2019). The authors were able to identify both positive and negative effects. On the one hand, it could be shown that leaderboards can have a motivating outcome if

they are used as a global standard of comparison and thus stimulate comparative behavior. On the other hand, it could be shown that leaderboards increased the perceived pressure of the subjects. Christy and Fox also examined the game design element leaderboard separately in a virtual reality based study (Christy & Fox, 2014). The authors show how the display of participants of different gender on a leaderboard affects the academic identification and performance of women. Differences were found between leaderboards of either male or female participants. Landers et al. (2017) examine the effect of the game design element in the context of goal setting theory. In this context, they examined how goals set with different difficulty in leaderboards affected the motivation of subjects. And another study with mixed results by Ortiz-Rojas et al. (2019) demonstrates the general need to investigate this individual game design element more closely. We agree with that and suggest well-developed research designs to investigate this and other game design elements.

One assumption made in many gamification studies is that different game design elements have different effects on basic psychological needs (Bräuer & Mazarakis, 2019; Sailer et al., 2013; van Roy & Zaman, 2019). Future research could address the different effects of individual game design elements on these basic psychological needs. This in turn may provide indications as to which game design elements should be combined to address all basic psychological needs and thus achieve the most motivating effect possible, considering the personality type of the subjects.

Although our study uses more game design elements than the average experiments analyzed by Hamari et al. (2014), still more game design elements might be beneficial (Groening & Binnewies, 2021), leaving aside whether an individual instead of a joint use of game design elements would not also be sufficient (Mazarakis, 2021). This can be seen as a research gap in our present study and thus as a possible outlook for further research. Ideally, it would be a platform where the experimental setting does not vary, but the number and quality of the different game design elements do.

A caveat is useful for understanding the feedback element of the game. In this study, a confirmational feedback was explicitly used as a game design element (Kapp, 2014b; Kapp et al., 2014; Mazarakis, 2015). Although one might assume that “everything is feedback” and, for example, a progress bar is thus also feedback, it is important to consider the scientific distinction. Werbach and Hunter (2012) differentiate three categories of game design elements in gamification: dynamics, mechanics and components (Werbach & Hunter, 2012, p. 78). Feedback is considered as a mechanic (Werbach & Hunter, 2012, p. 79), whereas, e.g., progression is considered as a dynamic (Werbach & Hunter, 2012, p. 78). It is not the purpose of this study to analyze the scientific nuances of the terminology. For the sake of completeness, however, it should be pointed out that it can be discussed whether feedback really has its justification for existence as an individual game design element, although the existing literature strongly assumes that it actually does (Geelan et al., 2015; Kapp, 2012, 2014a, 2014b; Kapp et al., 2014; Mazarakis, 2015; McGonigal, 2011).

Finally, the implementation of the narrative element can be overly complex compared to other game design elements. Developing an appropriate story that has the potential to motivate the user is a greater challenge than designing points, levels or progress bars (Prestopnik & Tang, 2015). Still, long-term effects might be superior for the game design element narrative, in contrast to easier to implement game design elements like points, badges or leaderboards. More research about the narrative element is necessary.

7. Conclusion

We analyzed the results of 505 subjects about the individual and combined effects of several different game design elements in a quiz, namely feedback, progress bar, badges, and narrative. The different game design elements vary in complexity, from rather simple ones like feedback to more elaborated ones like narrative. Also the elaborated selection takes into account established gamification elements like badges and feedback and less established elements like progress bar and narrative. We can show in this article that individual game design elements are sufficient to motivate individuals to answer questions in a quiz compared to a control group without gamification. Additionally, the game design elements progress bar and badges, always in combination with feedback, have motivated the subjects best to answer all questions. The findings can help to apply to similar quiz settings. For example, these can be a gamified online training course’s quiz within an enterprise (Stanculescu et al., 2016).

This study is an important milestone in the analysis of individual game design elements for gamification by applying a randomized between-group study design. We can back our findings with the experimental setting, where we could analyze each effect individually under the *ceteris paribus* clause. All results were achieved in the same setting without having to question internal or external validity due to different experimental settings or designs. We address the issue of insufficient empirical evidence of the impact of individual game design elements on the motivation of users. In this aspect, this article addresses an important research gap.

Our study has led to new conclusions in the field of gamification research regarding the effectiveness of individual and combined game design elements and giving support to all six hypotheses. In each condition the assumed effect vs. a control or comparison group was confirmed. The game design elements feedback, progress bar and narrative have in our experiment a significant motivating effect when used individually. In addition, a positive effect of combined game design elements could also be demonstrated, with best results for a combination of badges and feedback. In the individual application of the game design elements, the greatest effect was achieved with the narrative element.

Disclosure statement

No potential conflict of interest was reported by the authors.

Note

1. Indeed, a reference to the novel “The Hitchhiker’s Guide to the Galaxy” by Douglas Adams.

ORCID

Athanasios Mazarakis  <http://orcid.org/0000-0001-9943-0382>

Paula Bräuer  <http://orcid.org/0000-0001-5903-8829>

References

- Abramovich, S., Schunn, C., & Higashi, R. M. (2013). Are badges useful in education? It depends upon the type of badge and expertise of learner. *Educational Technology Research and Development*, 61(2), 217–232. <https://doi.org/10.1007/s11423-013-9289-2>
- Antin, J., & Churchill, E. F. (2011). Badges in social media: A social psychological perspective. In *CHI 2011 Gamification Workshop Proceedings* (pp. 1–4).
- Berger, E., Saethre, T. H., & Divitini, M. (2019). PrivaCity. In S. N. Pozdniakov & V. Dagienė (Eds.), *Proceedings of the 12th International Conference on Informatics in Schools: Situation, Evolution, and Perspectives, ISSEP 2019* (pp. 293–304). Springer International Publishing. https://doi.org/10.1007/978-3-030-33759-9_23
- Bevins, K. L., & Howard, C. D. (2018). Game mechanics and why they are employed: What we know about gamification so far. *Issues and Trends in Educational Technology*, 6(1), 58–84. https://doi.org/10.2458/azu_itet_v6i1_bevins
- Botra, A., Rerselman, M., & Ford, M. (2014). Gamification beyond badges. In *Proceedings of the 2014 IST-Africa Conference* (pp. 1–10). <https://doi.org/10.1109/ISTAFRICA.2014.6880651>
- Bräuer, P., & Mazarakis, A. (2019). Badges or a leaderboard? How to gamify an augmented reality warehouse setting. In J. Koivisto & Hamari (Eds.), *Proceedings of the 3rd International GamiFIN Conference—GamiFIN 2019* (Vol. 2359, pp. 229–240). CEUR-WS.org.
- Cheong, C., Cheong, F., & Filippou, J. (2013). Quick quiz: A gamified approach for enhancing learning. In *PACIS 2013 Proceedings*.
- Christy, K. R., & Fox, J. (2014). Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women’s math performance. *Computers & Education*, 78, 66–77. <https://doi.org/10.1016/j.compedu.2014.05.005>
- Conrad, F. G., Couper, M. P., Tourangeau, R., & Peytchev, A. (2010). The impact of progress indicators on task completion. *Interacting with Computers*, 22(5), 417–427. <https://doi.org/10.1016/j.intcom.2010.03.001>
- Csikszentmihalyi, M. (2009). *Flow: The psychology of optimal experience*. Harper Collins.
- Davidson, S., Candy, L. (2016). Gameful learning as a teaching/learning strategy: Best practices and lessons learned. In *Proceedings of the Edmedia + Innovate Learning 2016 Conference* (pp. 1165–1171). <https://www.learntechlib.org/primary/p/173093/>
- Deal, D. (2018, March 18). *Google Gamifies Google Maps*. Medium. <https://medium.com/@davidjdeal/google-gamifies-google-maps-81cfa0e4eedd>
- Dengri, C., Gill, A., Chopra, J., Dengri, C., Koritala, T., Khedr, A., Korsapati, A. R., Adhikari, R., Jain, S., Zec, S., Chand, M., Kashyap, R., Pattan, V., Khan, S. A., & Jain, N. K. (2021). A review of the quiz, as a new dimension in medical education. *Cureus*, 13(10), e18854. <https://doi.org/10.7759/cureus.18854>
- 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). <https://doi.org/10.1145/2181037.2181040>
- Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A critical review. *International Journal of Educational Technology in Higher Education*, 14(1), 9. <https://doi.org/10.1186/s41239-017-0042-5>
- Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Educational Technology & Society*, 18(3), 75–88. <https://www.jstor.org/stable/jeductechsoci.18.3.75>
- Domínguez, A., Saenz-de-Navarrete, J., de-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J.-J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education*, 63, 380–392. <https://doi.org/10.1016/j.compedu.2012.12.020>
- Epstein, J. A., & Harackiewicz, J. M. (1992). Winning is not enough: The effects of competition and achievement orientation on intrinsic interest. *Personality and Social Psychology Bulletin*, 18(2), 128–138. <https://doi.org/10.1177/0146167292182003>
- Facey-Shaw, L., Specht, M., van Rosmalen, P., Borner, D., & Bartley-Bryan, J. (2018). Educational functions and design of badge systems: A conceptual literature review. *IEEE Transactions on Learning Technologies*, 11(4), 536–544. <https://doi.org/10.1109/TLT.2017.2773508>
- Feger, S., Dallmeier-Tiessen, S., Woźniak, P., & Schmidt, A. (2018). Just not the usual workplace: Meaningful gamification in science. In *Mensch und Computer 2018 – Workshopband* (pp. 113–118). Gesellschaft Für Informatik e.V. <https://doi.org/10.18420/muc2018-ws03-0366>
- Feger, S., Dallmeier-Tiessen, S., Woźniak, P., & Schmidt, A. (2019). Gamification in science: A study of requirements in the context of reproducible research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems – CHI ’19* (pp. 1–14). <https://doi.org/10.1145/3290605.3300690>
- Field, A. P. (2009). *Discovering statistics using SPSS (and sex, drugs and Rock “N” Roll)* (3rd ed.). SAGE Publications.
- Finckenhagen, K. R. (2017). *Context in gamification—Contextual factors and successful gamification*. https://www.ntnu.edu/documents/139799/1279149990/09+Article+Final_karolirf_fors%C3%B8k_2017-12-08-05-35-05_TPD4505.Karoline.Finckenhagen.pdf/4f269657-0e54-4999-9de8-6ba8e109d386
- Friedrich, J., Becker, M., Kramer, F., Wirth, M., & Schneider, M. (2020). Incentive design and gamification for knowledge management. *Journal of Business Research*, 106, 341–352. <https://doi.org/10.1016/j.jbusres.2019.02.009>
- García-Marquez, C., & Bauer, K. N. (2021). An examination and extension of the theory of gamified learning: The moderating role of goal orientation. *Simulation & Gaming*, 52(4), 407–428. <https://doi.org/10.1177/1046878120958741>
- Geelan, B., De Salas, K., Lewis, I., King, C., Edwards, D., & O’Mara, A. (2015). Improving learning experiences through gamification: A case study. *Australian Educational Computing*, 30(1), 1–21. <http://journal.acce.edu.au/index.php/AEC/article/view/57>
- Gibson, D., Ostaszewski, N., Flintoff, K., Grant, S., & Knight, E. (2015). Digital badges in education. *Education and Information Technologies*, 20(2), 403–410. <https://doi.org/10.1007/s10639-013-9291-7>
- Grobely, J., Smierzchalska, J., & Czapkowski, K. (2018). Narrative gamification as a method of increasing sales performance: A field experimental study. *International Journal of Academic Research in Business and Social Sciences*, 8(3), 430–447. <https://doi.org/10.6007/IJARBSS/v8-i3/3940>
- Groening, C., & Binnewies, C. (2019). “Achievement unlocked!”—The impact of digital achievements as a gamification element on motivation and performance. *Computers in Human Behavior*, 97, 151–166. <https://doi.org/10.1016/j.chb.2019.02.026>
- Groening, C., & Binnewies, C. (2021). The More, the Merrier? – How adding and removing game design elements impact motivation and performance in a gamification environment. *International Journal of Human-Computer Interaction*, 37(12), 1130–1150. <https://doi.org/10.1080/10447318.2020.1870828>
- Hakulinen, L., Auvinen, T., & Korhonen, A. (2015). The effect of achievement badges on students’ behavior: An empirical study in a university-level computer science course. *International Journal of Emerging Technologies in Learning*, 10(1), 18. <https://doi.org/10.3991/ijet.v10i1.4221>
- 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 – CHI PLAY ’19* (pp. 559–572). <https://doi.org/10.1145/3311350.3347167>

- 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. <https://doi.org/10.1016/j.elerap.2013.01.004>
- Hamari, J. (2017). Do badges increase user activity? A field experiment on the effects of gamification. *Computers in Human Behavior*, 71, 469–478. <https://doi.org/10.1016/j.chb.2015.03.036>
- Hamari, J., & Koivisto, J. (2014). Measuring flow in gamification: Dispositional flow scale-2. *Computers in Human Behavior*, 40, 133–143. <https://doi.org/10.1016/j.chb.2014.07.048>
- Hamari, J., Koivisto, J., Sarsa, H. (2014). Does gamification work? – A literature review of empirical studies on gamification. In *47th Hawaii International Conference on System Sciences (HICSS)* (pp. 3025–3034).
- 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. <https://doi.org/10.1016/j.compedu.2014.08.019>
- Huschens, M., Rothlauf, F., Rothe, R. (2019). On the role of social comparison processes in gamified work situations. In *Proceedings of the 52nd Hawaii International Conference on System Sciences – HICSS 52* (pp. 1446–1455).
- Jia, Y., Xu, B., Karanam, Y., & Voids, 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). <https://doi.org/10.1145/2858036.2858515>
- 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. <https://doi.org/10.1016/j.invent.2016.10.002>
- Kalafatoglu, Y. (2020). Gamification in business: A review of the studies. In M. H. Bilgin, H. Danis, & E. Demir (Eds.), *Eurasian business perspectives* (Vol. 14, pp. 53–73). Springer. https://doi.org/10.1007/978-3-030-52294-0_4
- Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. John Wiley & Sons.
- Kapp, K. M. (2014a). Gamification: Separating fact from fiction. *Chief Learning Officer*, 13(3), 42–52. <https://www.chieflearningofficer.com/2014/03/19/gamification-separating-fact-from-fiction/>
- Kapp, K. M. (2014b, August 11). Game element: Feedback. *Kapp Notes*. <https://karlkapp.com/game-element-feedback/>
- Kapp, K. M., Blair, L., Mesch, R. (2014). *The gamification of learning and instruction fieldbook: Ideas into practice*. Wiley. <https://www.wiley.com/en-us/The+Gamification+of+Learning+and+Instruction+Fieldbook%3A+Ideas+into+Practice-p-9781118674437>
- Keusch, F., & Zhang, C. (2017). A review of issues in gamified surveys. *Social Science Computer Review*, 35(2), 147–166. <https://doi.org/10.1177/0894439315608451>
- Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, 45, 191–210. <https://doi.org/10.1016/j.ijim.2018.10.013>
- Kumar, J. (2013). Gamification at work: Designing engaging business software. In A. Marcus (Ed.), *Design, user experience, and usability: Health, learning, playing, cultural, and cross-cultural user experience* (pp. 528–537). Springer.
- Kywski, E., & Krämer, N. C. (2018). To gamify or not to gamify? An experimental field study of the influence of badges on motivation, activity, and performance in an online learning course. *Computers & Education*, 118, 25–37. <https://doi.org/10.1016/j.compedu.2017.11.006>
- 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. <https://doi.org/10.1177/1046878118774385>
- 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. <https://doi.org/10.1016/j.chb.2015.08.008>
- Liu, D., Santhanam, R., & Webster, J. (2017). Toward meaningful engagement: A framework for design and research of gamified information systems. *MIS Quarterly*, 41(4), 1011–1034. <https://doi.org/10.25300/MISQ/2017/41.4.01>
- Locke, E. (2001). Motivation by goal setting. In R. Golembiewski (Ed.), *Handbook of organizational behavior*. Dekker.
- Locke, E., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. *American Psychologist*, 57(9), 705–717. <https://doi.org/10.1037/0003-066X.57.9.705>
- Majuri, J., Koivisto, J., Hamari, J. (2018). Gamification of education and learning: A review of empirical literature. In *Proceedings of the 2nd International GamiFIN Conference, GamiFIN 2018* (pp. 11–19).
- Mazarakis, A. (2015). Using gamification for technology enhanced learning: The case of feedback mechanisms. *Bulletin of the IEEE Technical Committee on Learning Technology*, 4(17), 6–9. https://www.researchgate.net/publication/316384079_Using_gamification_for_technology_enhanced_learning_The_case_of_feedback_mechanisms
- Mazarakis, A. (2021). Gamification reloaded: Current and future trends in gamification science. *i-com*, 20(3), 279–294. <https://doi.org/10.1515/icom-2021-0025>
- Mazarakis, A., Bräuer, P. (2020a). Gamification of an open access quiz with badges and progress bars: An experimental study with scientists. In J. Koivisto, M. Bujic, & J. Hamari (Eds.), *Proceedings of the 4th International GamiFIN Conference—GamiFIN 2020* (Vol. 2637, pp. 62–71). CEUR-WS.org.
- Mazarakis, A., & Bräuer, P. (2020b). First directions for using gamification to motivate for open access. *arXiv 2002.03681 [Cs]*. <http://arxiv.org/abs/2002.03681>
- Mazarakis, A., & Bräuer, P. (2018). Gamification is working, but which one exactly? Results from an experiment with four game design elements. In *Proceedings of the Technology, Mind, and Society* (p. 22, 1). <https://doi.org/10.1145/3183654.3183667>
- McGonigal, J. (2011). *Reality is broken*. Penguin Press.
- Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, 71, 525–534. <https://doi.org/10.1016/j.chb.2015.08.048>
- 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). <https://doi.org/10.1145/2583008.2583017>
- Miller, A. S., Cafazzo, J. A., & Seto, E. (2016). A game plan: Gamification design principles in Mhealth applications for chronic disease management. *Health Informatics Journal*, 22(2), 184–193. <https://doi.org/10.1177/1460458214537511>
- Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (Vol. 2., pp. 745–783). Lawrence Erlbaum Associates.
- Muntean, C. I. (2011). Raising engagement in E-learning through gamification. In *Proceedings of the 6th International Conference on Virtual Learning (ICVL)* (pp. 323–329).
- Myers, B. A. (1985). The importance of percent-done progress indicators for computer-human interfaces. *ACM SIGCHI Bulletin*, 16(4), 11–17. <https://doi.org/10.1145/1165385.317459>
- Nah, F. F.-H., Telaprolu, V. R., Rallapalli, S., & Venkata, P. R. (2013). Gamification of education using computer games. In S. Yamamoto (Ed.), *Human interface and the management of information. Information and interaction for learning, culture, collaboration and business* (Vol. 8018, pp. 99–107). Springer. https://doi.org/10.1007/978-3-642-39226-9_12
- Ortiz-Rojas, M., Chiluiza, K., & Valcke, M. (2019). Gamification through leaderboards: An empirical study in engineering education. *Computer Applications in Engineering Education*, 27(4), 777–788. <https://doi.org/10.1002/cae.12116>
- Patchan, M. M., & Puranik, C. S. (2016). Using tablet computers to teach preschool children to write letters: Exploring the impact of

- extrinsic and intrinsic feedback. *Computers & Education*, 102, 128–137. <https://doi.org/10.1016/j.compedu.2016.07.007>
- Patzer, Y., Russler, N., & Pinkwart, N. (2018). Gamification in inclusive eLearning. In K. Miesenberger & G. Kouroupetroglou (Eds.), *Computers helping people with special needs* (Vol. 10896, pp. 154–158). Springer. https://doi.org/10.1007/978-3-319-94277-3_26
- Prestopnik, N. R., & Tang, J. (2015). Points, stories, worlds, and diegesis: Comparing player experiences in two citizen science games. *Computers in Human Behavior*, 52, 492–506. <https://doi.org/10.1016/j.chb.2015.05.051>
- Raikar, S. P. (2021, November 4). *Quiz*. Britannica. <https://www.britannica.com/topic/quiz>
- Ramaprasad, A. (1983). On the Definition of Feedback. *Behavioral Science*, 28(1), 4–13. <https://doi.org/10.1002/bs.3830280103>
- Rapp, A., Hopfgartner, F., Hamari, J., Linehan, C., & Cena, F. (2019). Strengthening gamification studies: Current trends and future opportunities of gamification research. *International Journal of Human-Computer Studies*, 127, 1–6. <https://doi.org/10.1016/j.ijhcs.2018.11.007>
- Richards, C., Thompson, C. W., & Graham, N. (2014). Beyond designing for motivation: The importance of context in gamification. In *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '14)* (pp. 217–226). <https://doi.org/10.1145/2658537.2658683>
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54–67. <https://doi.org/10.1006/ceps.1999.1020>
- 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. <https://doi.org/10.1007/s11031-006-9051-8>
- Sailer, M., Hense, J., Mandl, H., & Klevers, M. (2013). Psychological perspectives on motivation through gamification. *IXD&A*, 19, 28–37.
- Sailer, M., Hense, J., Mandl, H., & Klevers, M. (2017). Fostering development of work competencies and motivation via gamification. In M. Mulder (Ed.), *Competence-based vocational and professional education* (Vol. 23, pp. 795–818). Springer. https://doi.org/10.1007/978-3-319-41713-4_37
- 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. <https://doi.org/10.1016/j.chb.2016.12.033>
- Sanchez, D. R., Langer, M., & Kaur, R. (2020). Gamification in the classroom: Examining the impact of gamified quizzes on student learning. *Computers & Education*, 144, 103666. <https://doi.org/10.1016/j.compedu.2019.103666>
- Sardi, L., Idri, A., & Fernández-Alemán, J. L. (2017). A systematic review of gamification in E-health. *Journal of Biomedical Informatics*, 71, 31–48. <https://doi.org/10.1016/j.jbi.2017.05.011>
- Saxegaard, E., Divitini, M. (2019). ‘CITY: A Game to Raise Girls’ awareness about information technology. In S. N. Pozdniakov & V. Dagienė (Eds.), *Proceedings of the 12th International Conference on Informatics in Schools: Situation, Evolution, and Perspectives, ISSEP 2019* (pp. 268–280). Springer International Publishing. https://doi.org/10.1007/978-3-030-33759-9_21
- Schöbel, S., Söllner, M., & Leimeister, J. M. (2016). The agony of choice—Analyzing user preferences regarding gamification elements in learning management systems. In *International Conference on Information Systems (ICIS)*.
- Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74, 14–31. <https://doi.org/10.1016/j.ijhcs.2014.09.006>
- Sheffler, Z. J., Liu, D., & Curley, S. P. (2020). Ingredients for successful badges: Evidence from a field experiment in bike commuting. *European Journal of Information Systems*, 29(6), 688–616. <https://doi.org/10.1080/0960085X.2020.1808539>
- Siemens, J. C., Smith, S., Fisher, D., Thyroff, A., & Killian, G. (2015). Level up! The role of progress feedback type for encouraging intrinsic motivation and positive brand attitudes in public versus private gaming contexts. *Journal of Interactive Marketing*, 32, 1–12. <https://doi.org/10.1016/j.intmar.2015.07.001>
- Stanculescu, L. C., Bozzon, A., Sips, R.-J., & Houben, G.-J. (2016). Work and play: An experiment in enterprise gamification. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)* (pp. 346–358). <https://doi.org/10.1145/2818048.2820061>
- Sümer, M., & Aydın, C. H. (2018). Gamification in open and distance learning: A systematic review. In M. J. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, design, and technology* (pp. 1–16). Springer International Publishing. https://doi.org/10.1007/978-3-319-17727-4_115-1
- Tauer, J. M., & Harackiewicz, J. M. (1999). Winning isn’t everything: Competition, achievement orientation, and intrinsic motivation. *Journal of Experimental Social Psychology*, 35(3), 209–238. <https://doi.org/10.1006/jesp.1999.1383>
- Treiblmaier, H., Putz, L.-M., & Lowry, P. B. (2018). Setting a definition, context, and theory-based research agenda for the gamification of non-gaming applications. *Association for Information Systems Transactions on Human-Computer Interaction*, 10(3), 129–163. <https://doi.org/10.17705/1thci.00107>
- Vallerand, R. J., Gauvin, L. L., & Halliwell, W. R. (1986). Negative effects of competition on children’s intrinsic motivation. *The Journal of Social Psychology*, 126(5), 649–656. <https://doi.org/10.1080/0022454.1986.9713638>
- van Roy, R., Deterding, S., & Zaman, B. (2019). Collecting Pokémon or receiving rewards? How people functionalise badges in gamified online learning environments in the wild. *International Journal of Human-Computer Studies*, 127, 62–80. <https://doi.org/10.1016/j.ijhcs.2018.09.003>
- van Roy, R., & Zaman, B. (2019). Unravelling the ambivalent motivational power of gamification: A basic psychological needs perspective. *International Journal of Human-Computer Studies*, 127, 38–50. <https://doi.org/10.1016/j.ijhcs.2018.04.009>
- Vansteenkiste, M., & Deci, E. L. (2003). Competitively contingent rewards and intrinsic motivation: Can losers remain motivated? *Motivation and Emotion*, 27(4), 273–299. <https://doi.org/10.1023/A:1026259005264>
- Warmelink, H., Koivisto, J., Mayer, I., Vesa, M., Hamari, J. (2018). Gamification of the work floor: A literature review of gamifying production and logistics operations. In *Proceedings of the 51st Hawaii International Conference on System Sciences* (pp. 1108–1117).
- Warmelink, H., Koivisto, J., Mayer, I., Vesa, M., & Hamari, J. (2020). Gamification of production and logistics operations: Status quo and future directions. *Journal of Business Research*, 106, 331–340. <https://doi.org/10.1016/j.jbusres.2018.09.011>
- Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Wharton Digital Press.
- Xi, N., & Hamari, J. (2019). Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction. *International Journal of Information Management*, 46, 210–221. <https://doi.org/10.1016/j.ijinfomgt.2018.12.002>
- Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. O’Reilly Media.

About the Authors

Athanasios Mazarakis is a post-doc at the Computer Science department for Web Science at Kiel University and has been working on gamification and incentives in the interdisciplinary field between computer science, economics, and psychology for over a decade. Numerous publications on gamification and successful workshop organizations complete his competence profile.

Paula Bräuer is a research assistant and Ph.D. candidate at Kiel University since 2018. She has a master’s degree in business informatics and researches gamification in various contexts, such as augmented reality and intelligent virtual assistants.

CHAPTER

FOUR

Audio-Gamification Lab Experiment

In this chapter, the results of the laboratory study are presented in the form of a peer-reviewed journal article:

Bräuer, P., & Mazarakis, A. (2022). “Alexa, Can We Design Gamification Without a Screen?”—Implementing Cooperative and Competitive Audio-Gamification for Intelligent Virtual Assistants. *Computers in Human Behavior*, 135, 107362. <https://doi.org/10.1016/j.chb.2022.107362>

Impact Factor 2022: 8.9

The data were collected in the context of a seminar in 2019. The author and Dr. Athanasios Mazarakis supervised the seminar and jointly interrogated the study’s design. The implementation of the Alexa skill and the data collection were carried out by the student participants who attended the seminar. The analysis of the data and manuscript preparation was taken by the author of and revised by Dr. Athanasios Mazarakis.

4.1. Preface to Study 2

The first study of this thesis showed that individual game design elements can be sufficient for a successful implementation of gamification. This means that even simple incentives can motivate users. This insight is crucial for the implementation of gamification for IVAs because, without an additional screen, the possibilities to implement game design elements are pretty limited. In addition, the amount of information that can be delivered to a user at one time via audio is also limited (Bouzid & Ma, 2022). Therefore, it is crucial to investigate whether the integration of gamification

into IVA applications incentivizes their use or overloads the application and confuses the user. Therefore, the study presented in this chapter is the first to address whether it is possible to motivate users of IVAs by integrating game design elements. To answer this question, a laboratory experiment was conducted. The experiment compares two gamified groups with a control group without gamification. An Amazon Alexa skill was developed for this study to enable a comparison between different groups whose primary functionality is identical for all three versions. Different game design elements complement the functionalities in the two gamified versions. The decision to develop a skill rather than rely on a Wizard of OZ study design as in other studies is based on the need to create a setting that is as close to reality as possible. (McTear et al., 2016). In addition, using a skill ensured that Alexa's speech flow and length remained exactly the same to provide comparability between groups.

4.2. Technical Aspects of Study 2

In addition to the study results described in the journal article, this section provides insight into the technical implementation of the Alexa skill used.

For the lab study, a prototypical Alexa skill was developed that was run locally in developer mode and thus not publicly accessible. The flow of the study was designed to be linear so that the dialog between the subjects and the Alexa Speaker always followed the same pattern. Depending on which experimental condition a subject was randomly assigned to, a set of responses adapted for the condition was provided. Except for the wording of the responses, the procedure did not differ between experimental conditions.

Data logged as subjects interacted in the lab using the Alexa skill was stored as a JSON document in a NoSQL database. Each time the skill was started, a new entry was created in the database. The JSON document recorded how long it took subjects to complete each task. All other data was collected via an additional questionnaire that the subjects filled out after interacting with the skill.

4.3. Study 2 (publication)



“Alexa, can we design gamification without a screen?” - Implementing cooperative and competitive audio-gamification for intelligent virtual assistants

Paula Bräuer^{*}, Athanasios Mazarakis

Kiel University, Kiel, Germany

ARTICLE INFO

Keywords:

Gamification
Motivation
Performance
Intelligent virtual assistant
Amazon Alexa
Voice user interfaces

ABSTRACT

Intelligent virtual assistants (IVAs) like Amazon Alexa or Google Assistant have become increasingly popular in recent years, and research into the topic is growing accordingly. A major challenge in designing IVA applications is making them appealing. Gamification as a concept might help to boost motivation when using IVAs. Visual representation of progress and feedback is an essential component of gamification. When using IVAs, however, visual information is generally not available. To this end, this article reports the results of a lab experiment with 81 subjects describing how gamification, utilized entirely by audio, can assist subjects to work faster and improve motivation. Game design elements such as points and levels are integrated within an Alexa Skill via audio output to motivate subjects to complete household tasks. The results show a substantial effect on the subjects. Both their attitude and the processing time of the given tasks were positively influenced by the audio-gamification. The outcomes indicate that audio-gamification has a huge potential in the field of voice assistants. Differences in experimental conditions were also considered, but no statistical significance was found between the cooperative and competitive groups. Finally, we discuss how these insights affect IVA design principles and future research questions.

1. Introduction

Since the introduction of Siri for the Apple iPhone in 2011 (Guzman, 2019) intelligent virtual assistants (IVA) like Amazon Alexa, Google Assistant, or Microsoft Cortana have been introduced and are getting more and more popular (Clark et al., 2019; Dunn, 2016). In the literature, the term IVA is used interchangeably with terms like Intelligent Personal Assistant, Conversational Agent, Conversational User Interface, Virtual Personal Assistant, and Voice-Enabled Assistant, to mention a few (Cowan et al., 2017). These IVAs are operated primarily via voice and offer functions such as answering questions, playing music, controlling smart home devices, or creating shopping lists (Hoy, 2018). IVAs carry out about one billion tasks per month (Dellaert et al., 2020). Interaction with IVAs takes place via various interfaces, such as speaker devices, laptops, or smartphones (Guzman, 2019). In addition, wearables such as smartwatches and glasses¹ include an IVA, which can thus be used anytime and anywhere. To be able to react to user input, the speech recognition of the IVA is continuously active and waiting for

specific wake words, e.g., “Ok Google” or “Alexa.” If such a wake word is recognized, the voice input of the user is recorded and sent to a server for processing. Subsequently, the result of the request is sent back to the IVA, which transmits the (hopefully) desired information to the user via audio output (Hoy, 2018).

Often, speakers used to interact with an IVAs do not have a screen and can only reproduce information via audio output (Kinsella, 2019). This attribute has many advantages in everyday situations, such as driving a car, where the driver’s view should remain focused on the road (Hofmann, Tobisch, Ehrlich, Berton, & Mahr, 2014), or following a recipe in the kitchen when handling hot ingredients while cooking (Vtyurina & Fourney, 2018). IVAs are also increasingly being used to control augmented and virtual reality applications. For example, Cortana is one of the essential input tools for the augmented reality glasses HoloLens from Microsoft (Derby, Rickel, Harris, Lovell, & Chaparro, 2020). The IVAs offer a simple way of interacting with a system without the use of classic input tools such as the keyboard or partially unintuitive input via gesture control. This improves the immersive experience of

^{*} Corresponding author. ZBW/Kiel University, Düsternbrooker Weg 120, 24105, Kiel, Germany.

E-mail addresses: p.braeuer@zbw.eu (P. Bräuer), a.mazarakis@zbw.eu (A. Mazarakis).

¹ <https://www.amazon.com/Echo-Frames/dp/B07W72XKPJ>.

<https://doi.org/10.1016/j.chb.2022.107362>

Received 17 July 2021; Received in revised form 10 June 2022; Accepted 15 June 2022

Available online 18 June 2022

0747-5632/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

these applications (Shang & Wu, 2019).

IVAs also offer enormous potential in the healthcare sector. Interacting contactless with a computer can significantly reduce the risk of transmitting pathogens. For example, in sterile rooms, settings on devices can no longer be made at the touch of a button but simply by voice input (Mewes, Hensen, Wacker, & Hansen, 2017). It is also assumed that IVAs have enormous potential for e-marketing. Through AI-supported communication, for example, personalized consulting and shopping experiences can be created (Dellaert et al., 2020). However, the implementation of IVAs has not yet reached the point where users tend to personify them (Lopatovska & Williams, 2018). A study by Trajkova and Martin-Hammond (2020) showed that IVAs are currently rather poorly accepted by older adults for several reasons, including a lack of meaningful use cases or use in shared spaces, like, e.g., via cell phone while shopping.

Despite the increasing prevalence of IVAs and continued optimistic forecasts for the market (Shepherd & Liu, 2019), IVAs are often used by consumers only for a very limited set of tasks. Most of the time, IVAs are only used to play music or to ask about the weather (Lopatovska et al., 2019; Newman, 2018). In order to inspire users to use IVAs more widely, concepts are needed on how to make their use more interesting and motivating (Ji et al., 2017). As visual interaction via a screen is often missing, it is rather challenging to design motivational aids for IVAs because one can only rely on sound. In addition, a study by Montalvo et al. (2021) shows that interaction with an IVA is not considered easier when it has a screen, and usability is not improved compared to a device without a screen. Also, the effectiveness of performing tasks on IVAs with and without a screen does not seem to differ, so Montalvo et al. assume that an IVA without a screen is sufficient to meet the needs of most users. To motivate users to use an IVA and to make the applications more interesting, new approaches must be sought (Brandtzaeg & Følstad, 2018).

A common concept to engage and motivate users is gamification (Warmelink, Koivisto, Mayer, Vesa, & Hamari, 2018). Deterding, Dixon, Khaled, and Nacke (2011) define gamification as “the use of game design elements in non-game contexts.” To investigate the effect of game design elements on motivation and performance, among other aspects, studies have been conducted in various non-game contexts, such as work (Sailer, Hense, Mandl, & Klevers, 2017, pp. 795–818; Stanculescu, Bozzon, Sips, & Houben, 2016), health (Allam, Kostova, Nakamoto, & Schulz, 2015; Cechetti et al., 2019) and education (Dindar, Ren, & Järvenoja, 2021; Geelan et al., 2015). Most gamification studies also show a positive effect in many different contexts (Johnson et al., 2016; Majuri, Koivisto, & Hamari, 2018; Warmelink et al., 2018).

In Human-Computer Interaction (HCI), information has long been conveyed visually via graphical user interfaces. Acoustic output was only used to convey simple information such as alerts or in the area of multimedia applications (Sciarretta & Alimenti, 2021). Gamified applications also tend to present elements of game design visually. For example, the progress in a level is often displayed in the form of a progress bar (Zichermann & Cunningham, 2011), badges are explicitly distinguished by a visual representation (Antin & Churchill, 2011), and even just feedback (e.g., if an answer is right or wrong) is usually conveyed via visual elements such as checkmarks and crosses or colors, often using the colors red and green (Baneres, Rodriguez, & Serra, 2019; Czeszumski, Ehinger, Wahn, & König, 2019; Mazarakis & Bräuer, 2018, 2020, 2022). Implementing gamification for an IVA without visual support presents a vital challenge to overcome.

Not long ago, video games provided excitement and overcame the limitations of the graphics processing units, and visual representations using elaborated sound effects. Sound is used to highlight and support progress, feedback, and other game design elements (Collins, 2008). One prominent example is the game “Super Mario.” Each time the player collects a coin, i.e., scores a point, a characteristic sound is heard.

There are also plenty of purely audio-based computer games. These games offer, in particular for visually impaired people, the chance to

play video games, which is difficult to impossible with classic video games, as they put a strong focus on visual elements (Garcia & de Almeida Neris, 2013). However, since both the control (such as fast addressing correct items in a menu) and overview elements (e.g., a mini-map or the level display), as well as the actual activity and story, must be represented by sound, finding a suitable relationship between functionality and aesthetics poses a fundamental challenge in the development of audio games (Friberg & Gärdenfors, 2004). Similar challenges can also be found in the design of applications for IVAs, where it is currently often assumed that those aesthetics play a minor role in design (Harris, 2004).

However, this is a major challenge. So far, there is no scientific research on whether and how purely acoustically implemented game design elements can influence motivation to use IVAs. This article addresses this fundamental research gap and presents the results of an experiment on the use of game design elements presented only via audio output, which we define as *audio-gamification* (Mazarakis, 2021). Our pioneering study aims to provide insight into how to design and implement gamification, which relies only on audio feedback and without any visual representations.

The following section summarizes related work of virtual assistants and gamification as well as aspects regarding motivation, performance, cooperation, and competition. It is followed by the methods section, which presents our experiment. Chapter four shows the results as descriptive and inferential statistics. After discussing and interpreting our results, we conclude and present outlines for future work in the final chapter.

2. Theoretical background

Some of the basic principles on which this article is based include research on audio feedback, gamification, and IVAs. In the following, these areas are examined in more detail, particularly gamification and IVAs concerning the motivational context, performance aspects, and state of the art according to cooperation and competition. Finally, the hypotheses are derived based on this theoretical background.

2.1. Gamification and intelligent virtual assistants

Games for IVAs are a research area that has been almost ignored so far and where there is a need for further research (de Barcelos Silva et al., 2020). Though gamification is closer to game design than games (Landers, Auer, Collmus, & Armstrong, 2018; Mazarakis, 2021), there are many similarities, and contextual research is being performed on this paring (Leaning, 2015). By conducting research on gamification for IVAs, we can also contribute to addressing this research gap. Recently, the effect of game design elements in combination with an IVA got some attention, e.g., when learning English as a foreign language, which was investigated in an experiment by Tejedor-Garcia, Escudero-Mancebo, Cardenoso-Payo, and Gonzalez-Ferreras (2020). The authors describe the design and testing of a gamified application that can be used to practice the pronunciation of English vocabulary. The application was developed for smartphones and works with Google Assistant for voice recognition. To promote motivation, challenges were created in which users could compete. To investigate the effect of the game design elements on motivation and performance in pronunciation practice, a comparison was made with results from an earlier study (Tejedor-Garcia, Escudero-Mancebo, Gonzalez-Ferreras, Cámara-Arenas, & Cardenoso-Payo, 2016). The comparison showed positive trends in both learning success and motivation. For the implementation of the game design elements (challenge, leaderboard, points, avatar, and badges), visual representation was used. The results suggest that gamification is helpful in this context, but the question remains open whether gamification in combination with IVAs also works without visual components.

A very simple and, at the same time, very important element of gamification is feedback (McGonigal, 2011; Zichermann &

Cunningham, 2011). Feedback informs the user about their performance or the status of their actions, enabling behavior change (Kapp, 2012; Ramaprasad, 1983). Feedback can provide many functions like information, timing, error analyses, or in general for motivation (Mory, 2004). For example, participation can be increased by adding simple right/wrong feedback to a specific task (Mazarakis, 2015). So feedback can have different definitions and it is an overly complex concept (Ramaprasad, 1983). Therefore, to investigate the effect of gamification in the context of IVAs, audio feedback can be considered particularly important.

Thiebes, Lins, and Basten (2014) identified a comprehensive collection of 31 different game design elements through a systematic review of the game design elements proposed in the gamification literature. In their article, the authors list “audible feedback” as one of the game design elements they identified, which they separate from classic feedback. Audio feedback offers the possibility to create emotions and transmit information. This can be done via sound effects, other accompanying non-verbal effects, or music. Then again, it is also possible to use oral sound (or the spoken word), where information is given meaning through emphasis or rate of speech (Clark, 2003).

Speech is often used in audio games to illustrate and help in complex situations that are easier to explain verbally (Röber & Masuch, 2005). For example, in story-based adventure games, a narrator can develop the story and summarize scenes or events at certain points during the game. In audio games, speech is better suited for transmitting large amounts of information, whereas beacons and auditory textures perform better for simple and short messages (Röber & Masuch, 2005). In order for the user to know what the result of the input command is or whether the input was accepted, immediate feedback is required in purely acoustic applications (García & de Almeida Neris, 2013). For example, sound effects, such as footsteps while walking, show the player that a motion input was successful.

Research by Schmidmaier, Hußmann, and Runge (2020) deals with the effects of audio feedback on the formation of trust. The authors were able to experimentally demonstrate a significant effect of melody and mode on the formation of trust. This example shows how even simple adjustments can influence the effect of audio applications, such as skills for speech assistants.

Nykänen, Lopez, and Toulson (2016) investigated the effect of audio feedback on driving behavior. By using different forms of audio feedback, drivers should be motivated to drive as environmentally friendly as possible. Furthermore, the use of audio feedback should avoid the distraction of drivers that can occur when using a screen, for example. In a lab experiment, the authors were able to determine that signal tones and beep noises were received very differently while music or speech output was much better accepted. The results show that audio feedback is perceived very differently. It is already obvious that successful gamification with different audio game design elements is, therefore, a real challenge but at the same time also essential.

Silva-Coira, Cortiñas, and Pedreira (2016) establish a concept of how IVAs can be used in gamified environments. In the scenario presented, the IVA serves to help the user in their tasks by explaining the task or describing the use of tools. The goal of using IVAs in gamified environments is to detect the user’s mood using sentiment analysis and thus provide them with adapted answers. This should improve the experience of the application. Unfortunately, the authors describe the concept very superficial, and an evaluation of the concept is missing.

Ji et al. (2017) conducted a study with more than 20.000 users of an Alexa skill to identify ways to maintain the user’s motivation to use the IVA. To test the effect of different design variants, the IVA was extended with different modules during the test period. Subsequently, the trend of the rating of the skill on the Amazon platform was set in relation to the adjustments made. The authors found that, in general, a variety of conversational activities is desirable, and users tend to prefer natural conversations over menu-based conversations. It was also shown that the length of interactions with the conversation-based modules

corresponded positively with user ratings. The user evaluation was also positively influenced by the inclusion of narratives.

Stories or narratives are used in many games, and gamification research has already considered this element (Grobelny, Smierzchalska, & Czapkowski, 2018; Mazarakis & Bräuer, 2018, 2022; Sailer, Hense, Mayr, & Mandl, 2017, pp. 795–818). This game design element is particularly suitable for gamifying applications for IVAs, as it can be implemented well without a visual component. Narratives are often used in audio games for the same reason (Röber & Masuch, 2005). Based on a detailed literature review, Keusch and Zhang (2017) point out that it is important for the impact of a narrative integrated into an application to be adapted to the context. Therefore, despite the promising potential of narratives in the context of IVAs, attention should be paid to the design of the story for the respective application.

Landesberger, Ehrlich, and Minker (2020) also recognized the potential of IVAs in cars. They conducted a gamified study to investigate how subjects react to urgent acoustic messages during an activity, such as driving. Unfortunately, the study did not use a purely acoustic application for this purpose but developed a kind of visual game that was interrupted by acoustic subtasks. Information on progress and feedback on the completed tasks was reflected in the study by visual elements. Also, Fadhil and Villafiorita (2017) use a visual representation of game design elements in their study, in which they investigate the effect of gamification on learning with the help of conversational user interfaces. Since the chatbot in their study is a text-based language assistant, a visual representation is appropriate, but this does not apply to purely language-based agents. However, to obtain clear insights into the effect of audio-gamification in situations where no screen is available and therefore is not influenced by visual elements, it is necessary to conduct pioneering research.

2.1.1. Motivation aspects

The relevance of user motivation in human–artificial intelligence interactions is becoming increasingly apparent as the usefulness of today’s IVAs grows (Li & Yanagisawa, 2021). An initial work addresses what positive and negative experiences can be identified when using IVAs and how they affect the motivation to use them (Weber & Ludwig, 2020). The goal of gamification is usually to increase the motivation of the users for a given task or activity (Warmelink, Koivisto, Mayer, Vesa, & Hamari, 2020). Motivation in this context is leaned on Deci and Ryan’s self-determination theory (SDT) (Ryan & Deci, 2000). This motivational theory is the most popular in gamification research (Seaborn & Fels, 2015).

According to SDT, motivation can range between extrinsic and intrinsic motivation. Intrinsic motivation is defined as doing something because it is inherently interesting or fun. On the other hand, extrinsic motivation refers to doing something because it leads to a definable result (Ryan & Deci, 2000, 2017). A derived theory of SDT is the theory of basic psychological needs (Deci & Ryan, 2000). This theory argues that there are three basic psychological needs: experience of competence, sense of autonomy, and social relatedness. Experience of competence relates to a need for challenge and feelings of effectance (Deci & Ryan, 2000). Autonomy describes a sense of volition or willingness when doing a task (Deci & Ryan, 2000), and social relatedness is said to occur when a person feels connected to others (Guardia, J. G., Ryan, Couchman, & Deci, 2000). Fulfilling these three needs can give rise to intrinsic motivation. The theory assumes that the three basic psychological needs are influenced by an individual’s environment. This influence can be both positive and negative and can contribute to satisfying the needs or inhibiting their fulfillment (Vansteenkiste & Ryan, 2013).

Computer game enjoyment has also been positively related to fulfilling those needs (Pe-Than, Goh, & Lee, 2014; Przybylski, Rigby, & Ryan, 2010). With multiple experiments, Ryan, Rigby, and Przybylski (2006) demonstrated that when playing alone, a relationship between autonomy and competence satisfaction could be demonstrated, as well

as a relationship between all three needs in multiplayer contexts.

Many studies have shown that gamification influences motivation (Hamari, Koivisto, & Sarsa, 2014; Seaborn & Fels, 2015). Sailer, Hense, Mayr, and Mandl (2017), for example, looked at the effect of different game design elements on the three basic needs. The results of their study show that perceptions of competence can be positively influenced by badges, leaderboards, and performance graphs. Furthermore, avatars and narratives have shown a positive effect on the perception of social inclusion. Although most studies investigating the impact of gamification on motivation have yielded positive results, the context in which the game design elements are used has been shown to influence the effect (Hamari et al., 2014).

Research on IVAs has also recognized SDT as a relevant research concept. For example, Li and Yanagisawa (2021) showed that reducing uncertainty can lead to more interactions with an IVA, thus shifting the motivation for the interactions from non-intrinsic to intrinsic. Here, the authors assume that intrinsic motivation is present in the context of IVAs when activities are performed to pass the time and for play activities. Extrinsically motivated interactions, on the other hand, are pragmatically motivated to obtain information or to have certain tasks done by the assistants to save time.

Another important theory in terms of motivation and gamification is the flow theory. Flow theory should be considered in the context of intrinsic motivation, as the theory consists of aspects like immediate feedback and pursuing an activity for its own sake (Csikszentmihalyi, 2009). Feedback is crucial because an individual can get the necessary information on whether the activity in question is performed as intended. Furthermore, immediate and constant feedback helps to maintain the flow experience (Kapp, 2012). Based on the current state of research on the motivational effects of gamification and the lack of studies examining this very effect through audio-gamification, the following general research question is posed as a basis for upcoming hypotheses:

RQ1. How does audio-gamification affect motivation?

2.1.2. Performance aspects

In addition to the effect of gamification on motivation, studies also examine how gamification affects performance (Groening & Binnewies, 2019; Landers, Bauer, & Callan, 2015; Mekler, Bruehlmann, Tuch, & Opwis, 2017). Depending on the setting, a change in performance can address various metrics, such as quality, quantity, speed, or accuracy. For example, this can be beneficial in the work environment (Warmelink et al., 2018, 2020) when workers perform gamified tasks faster or more accurately or in learning (Bai, Hew, & Huang, 2020; Sailer & Homner, 2020), when students spend more time on a task and better results are achieved.

Groening and Binnewies (2019) show that individual game design elements can influence performance using the example of achievements. In a study with a within-subjects design, in which images were annotated, it could be shown that performance could be increased by adding achievements. Likewise, Mekler et al. (2017) demonstrated positive results by adding points, levels, and leaderboards in a similar setting. In a field experiment, Grobelny et al. (2018) demonstrated a positive effect on sales performance with the help of a narrative-based gamification approach. For this purpose, the sales performance of two franchisees was recorded and compared over two months. Moreover, in logistics, Sailer, Hense, Mandl, and Klevers (2017) were also able to show in an experimental study that performance in order picking can be increased through gamification. Bräuer and Mazarakis (2019) could support these findings in a different logistics study and show that some game design elements might be unsuitable in a specific context to enhance performance.

Based on this body of research, we assume that performance can also be influenced by purely audio-based game design elements, and the following research question was posed:

RQ2. How does audio-gamification affect performance?

2.1.3. Aspects on cooperation and competition

Since numerous studies have already proven the effect of gamification (Hamari et al., 2014; Mekler et al., 2017; Sailer, Hense, Mayr, & Mandl, 2017, pp. 795–818), current research is also concerned with investigating which different design approaches are particularly effective in the implementation of gamification (Rapp, Hopfgartner, Hamari, Linehan, & Cena, 2019). An important difference when implementing a game design element is whether the element is designed cooperatively or competitively. In competitive games, such as car racing, the objective is to compete with one or more opponents and win against them. It is about being faster, smarter, or simply more skilled than the opponents (Kapp, 2012). Cooperative design is about working together to achieve a common goal or result. This concept is often found in app games for smart devices, such as Candy Crush or FarmVille. By cooperating with friends and exchanging or giving virtual goods, the progress of the individual players can be accelerated significantly. They are encouraged not to work alone towards a specific goal but to team up with other players (Kapp, 2012).

How competition affects performance in a gamified setting was, e.g., experimentally investigated by Landers, Collmus, and Williams (2019) in the context of brainstorming. In comparison to a control group, it was shown that both creativity and the quality of the brainstorming contributions could be improved by competition. The authors assume that the motivational effect of the competition must have come about either through unconscious influence or through the implicit creation of extrinsic rewards for increased effort (Landers et al., 2019).

Morschheuser, Hamari, and Maedche (2019) conducted a study to discover the differences between cooperative, competitive, and inter-team competitive gamification. In a field experiment in a crowdsourcing context, the authors demonstrated that inter-team competition had the most significant positive effect on the participation and enjoyment of participants. However, a different study that examined the effect of cooperative and competitive gamification on the flow perception of students did not find differences between the different design approaches (Marinho et al., 2019). Nevertheless, the study results are limited due to the data of only 18 individuals.

Dindar et al. (2021) investigated the effects of cooperative and competitive gamification with the help of a gamified app for learning vocabulary. The comparison between two groups, one cooperative and one competitive, did not reveal any differences in the subjects' learning success, performance, or motivation. Unfortunately, no control group without gamification was used in their study. Thus, it cannot be ruled out that the overall gamification design may have achieved the desired effect. The authors claim that "The current findings emphasize that the positive influence of gamified cooperation on creating meaningful connections amongst learners should not be ignored, even though it facilitates similar learning and motivational outcomes as gamified competition." (Dindar et al., 2021, p. 142). This is contrary to the claims of, e.g., Morschheuser et al. (2019) and Landers et al. (2019).

Although the research gap of cooperative and competitive gamification has been identified and initial studies on this topic exist, the results are not always conclusive, and their interpretation is not always clear. Furthermore, there is a lack of studies that compare the effect of cooperative and competitive gamification in a controlled setting. Therefore, to close this research gap and test whether the advantages of cooperative team competition can be transferred to the context of audio-gamification, RQ3 and RQ4 were formed.

RQ3. How do cooperation and competition influence motivation when using audio-gamification?

RQ4. How do cooperation and competition influence performance when using audio-gamification?

2.2. Hypotheses

So far, no studies have explicitly addressed audio-gamification and

its effects on motivation and task performance. The research goal of this pioneering study is to investigate how audio-only implemented game design elements can be deployed successfully and how they can enhance motivation and performance. Six hypotheses were established to systematically investigate this research gap and to answer all four research questions presented in this chapter.

Hypotheses **H1a** and **H1b** address **RQ2**, the relationship between audio-gamification and performance. We argue that, based on the results of [Montalvo et al. \(2021\)](#), task performance when using an IVA is not negatively affected by the absence of a screen. At the same time, we assume that the positive effect on performance induced by primarily exclusively visual gamification can also be achieved by audio-gamification, similar to audio games.

H1a. Subjects in the experimental condition with *competitive* audio-gamification perform tasks significantly faster than the control group.

H1b. Subjects in the experimental condition with *cooperative* audio-gamification perform tasks significantly faster than the control group.

Hypotheses **H2a** and **H2b** address **RQ1**, namely the relationship between audio-gamification and motivation. According to our assumptions when developing the hypotheses about performance, we believe that the numerous beneficial results obtained via gamification on motivation can also be obtained through audio-gamification.

H2a. Subjects in the experimental condition with *competitive* audio-gamification feel significantly more motivated than the control group.

H2b. Subjects in the experimental condition with *cooperative* audio-gamification feel significantly more motivated than the control group.

The last two hypotheses address the difference between cooperatively and competitively implemented audio-gamification. Hypothesis **H3** addresses **RQ3**, and **RQ4** is handled by hypothesis **H4**.

H3. Subjects in the experimental condition with *cooperative* audio-gamification perform tasks significantly faster than the subjects in the experimental condition with *competitive* audio-gamification.

H4. Subjects in the experimental condition with *cooperative* audio-gamification feel more motivated than the subjects in the experimental condition with *competitive* audio-gamification.

3. Method and data

To investigate the effects that audio-gamification can have on motivation and performance, a lab experiment was conducted. The following section describes the experimental design, the operationalization of the dependent and independent variables, and descriptive data about the subjects of the study in detail.

3.1. Experimental design

We created for our experiment a skill for the IVA Amazon Alexa to test in a lab experiment. A skill is similar to an app for a smartphone ([Zhang et al., 2019](#)). The skill extends the range of actions that can be performed by the IVA, using voice control provided by Amazon via a corresponding device, e.g., Amazon Echo.

The study uses a 3×2 between-subject design. The subjects are randomly assigned to one of three groups. One group serves as a control group without any game design elements. In contrast, the other two experimental groups (competition group and cooperation group) use audio-implemented game design elements, both with five different game design elements: points, time pressure, level, narrative, and ranking. The groups are compared after the experiment to identify possible differences in the performance and motivation of the subjects to find support for our hypotheses, including data from a subsequent questionnaire administered immediately after the lab experiment.

Similar to the study by [Silva-Coira et al. \(2016\)](#), the IVA supports the user in performing tasks. The tasks and the general course of the

experiment do not differ between the groups. Only the subjects in the experimental groups are given additional audio game design elements built into the skill. In the following, we will first describe the general design of the experiment as it was implemented for the control group. Then we will describe the different game design elements and how they were implemented in the two experimental conditions.

3.1.1. Control group

Each subject was greeted and briefed that the purpose of the experiment was to evaluate a new skill. The subjects are not aware that there is more than one test condition. First, a tutorial familiarized the subjects with the use of Alexa and the skill. They learned which commands they had to use, for example, to start new tasks or how to get information about the current task. Each subject was then taken to a separate room where there was an Alexa Echo Dot speaker, and the subjects did start with the experiment. During the whole experiment, a supervisor was present in the same room. The supervisor was available to help in the unlikely event that a subject has problems interacting with the IVA or to help them with any technical or other issues.

The IVA explains six different tasks to the subjects during the experiment, one by one. The tasks should be reminiscent of daily household chores, which are usually done rather reluctantly. The tasks include activities such as cleaning the room or sorting books into a shelf. To receive new tasks or to confirm when a task has been completed, the subject interacts with the Alexa Skill. By saying, "I'm done!" the subject confirms that he or she has successfully completed the last task. [Fig. 1](#) shows an example dialogue of a subject in the control group.

Because in a real-world setting usually, it is not possible to check whether the tasks have actually been finished completely and correctly, the supervisor also did not intervene in our laboratory setting. We expected that the mere presence of a supervisor will not lead to cheating. If a subject deviates from the task, the supervisor of the experiment will note this silently to ensure a high internal validity of the experiment.

After the completion of all tasks, a questionnaire had to be filled out to obtain information about the motivation during the activity as well as feedback on the audio implementation. The subjects in the experimental groups are also asked whether and how they perceived the individual game design elements. Since other studies have already found differences in the effect of gamification on gender ([Codish & Ravid, 2017](#); [Koivisto & Hamari, 2014](#)) and age ([Jent & Janneck, 2018](#)), this information is also collected. We provide additional information in section [3.2](#) about the questionnaire.

Next, we provide a short overview of the different game design elements and their implementation in the experiment. The basic structure of both experimental groups corresponds to the structure of the control group and differs only in these game design elements.

3.1.2. Points

Points are the basis of most gamified applications ([Zichermann & Cunningham, 2011](#)) and offer a possibility to show the progress of the users in numerical form ([Seaborn & Fels, 2015](#)). Subjects in the gamified experimental groups get for each successfully accomplished task 50 points. In addition, extra points can be earned depending on the speed performance. A maximum of 25 bonus points is awarded if the subject completes a task within the minimum duration, which was previously determined based on a pre-test. Therefore, a maximum of 75 points per task can be obtained, which makes a maximum of 450 points for the complete experiment (six tasks).

The subject will be informed about how to earn points and bonus points when the skill is started. This description and the number of points achieved are always communicated to the subject exclusively by audio through the Alexa Speaker. An example of a dialogue with a gamified version of the skill is given in [Fig. 2](#).

3.1.3. Time pressure

By awarding the extra points and announcing them at the end of a



Fig. 1. Example of a dialogue of a subject in the control group.



Fig. 2. Example of a dialogue between a subject and a gamified version of the Alexa skills.

task, time pressure is created in this study. In many games, this game design element is used, often in a visual form, e.g., by a clearing hour-glass or displaying a countdown. By communicating time pressure, the stress level of the players is increased, and the motivation to act is stimulated (Kapp, 2012).

Time pressure can be used to create a form of challenge. However, it is not always helpful, especially when tasks need to be completed precisely (Thiebes et al., 2014). For this reason, time is only measured here for the awarding of extra points.

3.1.4. Level

Levels can have a similar effect as given goals and motivate a subject to achieve them (Koivisto & Hamari, 2019). In games, they can display ranks that the player can reach or refer to stages or areas in a game world (Seaborn & Fels, 2015). Often, progress bars are used to show how far a player has progressed (Zichermann & Cunningham, 2011).

If a certain number of tasks are completed in our studies, the subject is informed via sound output that a new level has been reached. The IVA says: "Yippee, you've reached level two!". A total of three levels were provided. The levels are firmly linked to the completion of the tasks. The subject moves up a level for the first time by completing the first three tasks. The second level is completed as soon as the fifth task is completed. The third and last level ends with the last task. As a result, we used levels to inform the subjects about their progress.

3.1.5. Narrative

In addition to points, time pressure, and levels, the game design element *narrative* is integrated into the skill. If a user earns points, the story unfolds gradually. This form of gamification is reminiscent of one of the first forms of computer games, so-called text-adventures (Atkinson, Baier, Copplestone, Devlin, & Swan, 2019). In this type of game, the player goes through a story that is told exclusively through text output. Also, any interaction of the player takes place via text input. A similar principle was applied in this study to design the narrative element. By interacting with the IVA via speech, the subject progresses through the story. Since Ji et al. (2017) showed that users prefer natural conversations to menu-based navigation, it was expected that the integration of the narrative would have a positive impact on the perception of the skill.

To create the most immersive experience possible with the IVA, an attempt was made to choose the narrative to match the context as much as possible (Keusch & Zhang, 2017). The narrative that was developed for this skill is loosely based on the video game "Portal" ("Portal," 2019). The subjects interact with an artificial intelligence called "TILA." TILA is rather grumpy in the beginning but will become nicer the more tasks are

fulfilled. While the user is working on the tasks, TILA tells anecdotes and facts that are vaguely related to the content of the current task. When cleaning up, for example, she curses about the pollution of the environment by humans.

Throughout the story, the subject notices that TILA gives the subject tasks to help her get a password to escape from the speaker. The intermediate dialogs not only help develop the narrative but also assist from a technical perspective. The Alexa Skills Kit limits the maximum break time a skill remains active when there is no user input to 10 s (Speech Synthesis Markup Language (SSML) Reference | Alexa Skills Kit, n.d.). After this period, the speaker shuts down and must be restarted by the wake word. The intermediate dialogs thus help to extend the time limit of 10 s.

3.1.6. Ranking

A social component can be created by comparing the results with others using rankings or leaderboards (Kapp, 2012). Rankings are therefore also a popular element in gamification research for generating cooperation or competition (Sailer, Hense, Mayr, & Mandl, 2017, pp. 795–818). In this study, the game design element ranking was implemented in two different conditions, cooperative and competitive. An overview of how the ranking was implemented in both gamified experimental conditions is shown in Fig. 3.

In the cooperative group, the subjects are told at the beginning of the story that they are part of Team Bravo and that this team competes against the fictitious Team Alpha. When the intermediate ranking is announced, subjects in the cooperative group receive information on how their Team Bravo compares with the fictitious Team Alpha. The information on how their team compares with the Alpha Team is given four times: after completing the second, third, fifth, and sixth task. It was decided not to do this after each task since it would be too monotonous in combination with the regular announcement of the points, and TILA would have talked too much, maybe leading to distraction.

Subjects in the competitive group do not receive any information at the beginning that a comparison with other people is taking place. Just after the second task has been completed, the subjects are informed whether they performed better, equally well, or worse than the previous subjects. It was decided that the design of the ranking should be based on a general statement (see Fig. 3) and not on a list of best scores with exact rankings. The reason for this decision was that it is much easier to understand instead of a leaderboard where several users and scores have to be listed. The competitive ranking is announced at the same four stages as in the cooperative group.

Whether a subject performs well, moderately, or poorly in the

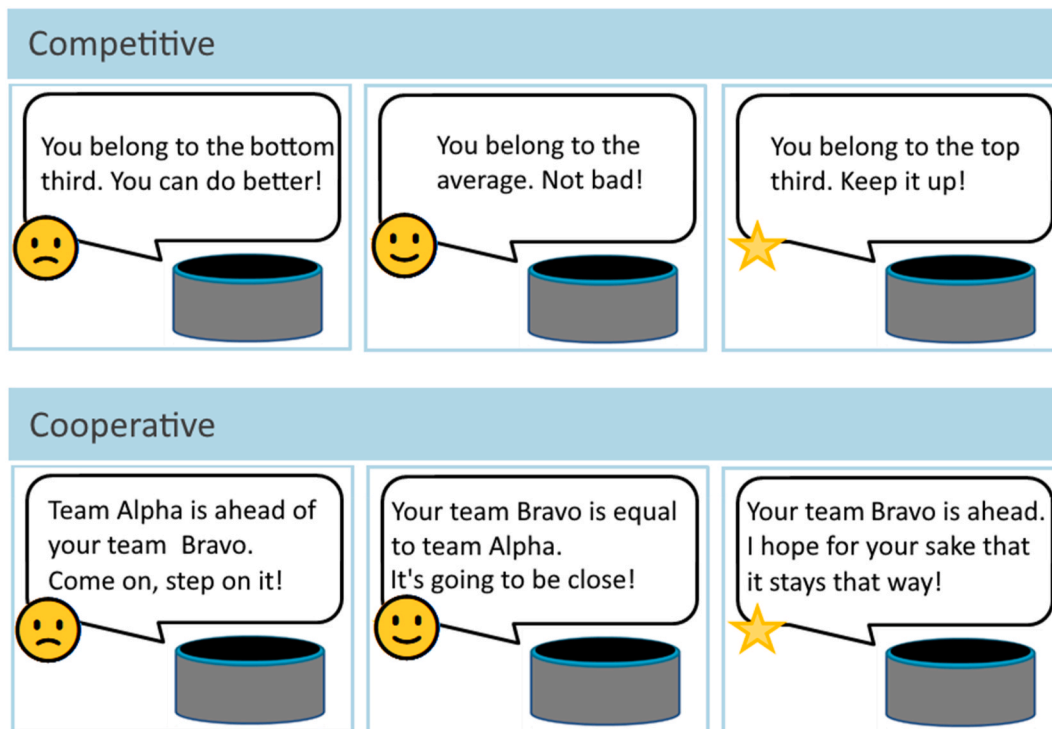


Fig. 3. Description of the cooperative and competitive ranking in the three gradations negative, neutral, and positive.

ranking is determined for each task based on the time it takes to complete the task. Therefore, the data for assessment on which the ranking is based is the same for all subjects and, like the points, is based on empirical data that has been collected by testing the tasks several times in a pre-test.

3.1.7. Audio feedback

In addition to the five game design elements mentioned, audio feedback is used to make the application more appealing. According to Thiebes et al. (2014), audio feedback includes both sound effects and music. We used short positive and negative mood audio files to provide positive or negative feedback to subjects. Furthermore, we used sound effects like a drum roll to create excitement. A combination of sound effects and music can also be found in audiobooks, for example, where they are used to create mood and atmosphere (Steinhaeuser, Schaper, & Lugrin, 2021). The audio files of the study supported the narrative and helped to give feedback in a simpler and more concise way than it would have been possible through detailed verbal responses (Röber & Masuch, 2005).

3.2. Measurements

In the study, the subjects' task completion time and motivation were analyzed. To evaluate the task completion time, the time required by the subjects to perform the individual tasks was measured. To be able to compare the times of the different test conditions, the exact completion time was collected. For the analysis, this completion time was adjusted by the time TILA was talking. For the final analysis, the total adjusted time on task for each individual was used.

An additional survey was immediately conducted after the laboratory experiment. The subjects had to answer a short questionnaire via a Google Forms sheet. In addition to demographic data (age and sex), the subjects were also asked about their previous usage behavior and experience with IVAs. The subjects could choose between six possible options ("never," "occasionally," "several times a month," "several times a week," "daily," and "several times a day"). Additionally, the subjects

were asked to indicate how motivated they were during the experiment. For this purpose, the following three questions were asked, which were evaluated based on a 5-point Likert scale (1 = not at all; 5 = very much):

- Q1 - Were you motivated by the virtual assistant (TILA) to solve the tasks?
- Q2 - Were you motivated by the virtual assistant (TILA) to solve the individual tasks faster?
- Q3 - Did you have fun with the virtual assistant (TILA)?

To perform a manipulation check, we asked the subjects in the questionnaire whether and if so which game design elements they perceived during the study. The subjects could first choose whether they thought they had perceived game design elements and then tick which of the five game design elements (points, time pressure, level, narrative, and ranking) they had perceived.

3.3. Participants

Data were collected from 91 participants at a German university. Of these, ten had to be excluded from evaluation because of incomplete data or technical or other issues. For example, contradictory results were obtained due to Internet failures or because subjects skipped entire tasks. Of these ten excluded subjects, four were in the control group, three in the group with cooperative gamification, and three in the group with competitive gamification. In total, a final sample size of $N = 81$ is obtained for further analysis.

The average age of the sample was 23.56 years (SD 4.42), with the youngest subject 17 years old and the oldest 41 years. 53 subjects were male and 28 female. The high proportion of students (85%) is given that the acquisition of the subjects was carried out at the university. The other subjects were employees.

Concerning the adoption and frequency of use of IVAs, 57 of the subjects (70.4%) stated that they had never used IVAs. Compared to the average use of IVAs, 26.2% in the U.S. and 25.8% in Germany (Kinsella, 2021), the proportion of people who at least occasionally use an IVA is

29.6% in our sample, which is comparable. An overview of the exact distribution in our study is given in Table 1. In addition, comparing the usage with a pre-COVID-19 survey with more than 1660 adults in the U.S.A., similarities can be found for weekly and monthly usage behavior (NPR and Edison Research, 2020).

4. Results

In this section, descriptive statistics are first briefly presented. Subsequently, the hypotheses are evaluated in three stages: first, the effect of audio-gamification on performance is examined, followed by the effect on motivation and the comparison between cooperative and competitive implementation. Finally, we consider whether there is an influence of variables such as age, gender, or previous experience with IVAs on the variables investigated.

Table 2 gives an overview of the number of subjects in the three test conditions, as well as the mean and median value and standard deviation of the time in seconds required by the subjects to complete all tasks. The unequal number of subjects per test condition is due to the random distribution of the subjects. Overall, the mean value for the adjusted required time is 544.78 s with a standard deviation of 115.77 s, and the median is 547.34 s. The fastest subject is from the cooperative condition with 331.58 s, and the slowest is also a subject from the same group with 838.32 s.

4.1. Effect on performance

To evaluate the hypotheses, it is first checked whether the collected data are normally distributed. For the hypotheses, H1a and H1b (subjects in experimental conditions are faster than the control group) as well as H3 (the cooperative group is faster than the competitive group), the normal distribution of the adjusted required time is tested first. According to the Shapiro-Wilk test (Field, 2017), no statistically significant result is provided, $W(81) = 0.98, p = .308$, and therefore a normal distribution of the data can be assumed.

Then the variance homogeneity of the sample is checked. The Levene test (Field, 2017) does not provide a statistically significant result, $F(2, 78) = 0.07, p = .936$. Thus, it can be assumed that the variance of the mean values for all three groups is not different, and a *t*-test can be performed to evaluate the hypotheses H1a, H1b, and H3.

Already from the descriptive results in Table 2, it can be seen that the subjects from the competitive group completed the tasks faster than those in the control group. Furthermore, the statistical comparison of the mean values between the competitive group ($M = 490.37, SD = 99.31$) and the control group ($M = 635.57, SD = 94.00$) also provides a statistically significant result with a large effect, $t(46) = 5.17, p = .000, d = 1.50$. Thus, H1a can be supported, and it is assumed that integrating audio-gamification with competitive design can positively influence performance.

The cooperative group's descriptive results from Table 2 also show that the tasks were completed faster than in the control group. Furthermore, the mean value comparison between the cooperative group ($M = 527.11, SD = 107.73$) and the control group ($M = 635.57, SD = 94.00$) is also statistically significant with a large effect, $t(53) = 3.84, p = .000, d = 1.07$. Thus, H1b can also be supported, and it

Table 1
Overview of the degree of previous experience with IVAs of the subjects.

| Frequency of use | N | Percentage of active users |
|-----------------------|----|----------------------------|
| Never | 57 | – |
| Occasionally | 14 | 58.33% |
| Several times a month | 4 | 16.67% |
| Several times a week | 2 | 8.33% |
| Daily | 4 | 16.67% |
| Serval times a day | 0 | 0% |

Table 2

Number of subjects, as well as the mean, median, and standard deviation of the adjusted required time per experimental group.

| Condition | N | M | Mdn | SD |
|-------------|----|--------|--------|--------|
| Control | 22 | 635.57 | 607.79 | 94.00 |
| Competitive | 26 | 490.37 | 504.80 | 99.31 |
| Cooperative | 33 | 527.11 | 526.48 | 107.73 |

can be assumed that the inclusion of both competitive and cooperative audio-gamification will positively impact performance.

4.2. Effect on motivation

The normal distribution of the questionnaire items Q1, Q2, and Q3 is checked first to evaluate the hypotheses H2a and H2b (subjects in the experimental conditions are more motivated than the control group) and H4 (the cooperative group is more motivated than the competitive group). The Shapiro-Wilk test provides a significant result for all three items: Q1 - *General Motivation* $W(81) = 0.89, p = .000$; Q2 - *Motivation to work fast* $W(81) = 0.90, p = .000$; Q3 - *Feeling of fun* $W(81) = 0.87, p = .000$. The data of all three items are therefore not normally distributed, and a non-parametric method (Mann-Whitney-U-Test) is used for further evaluation.

The results of the mean value comparisons for the evaluation of H2a are summarized in Table 4, the corresponding median values can be found in Table 3. For all three items, the Mann-Whitney U Test provides a statistically significant result. Furthermore, a medium effect was determined for the items Q1 - *General Motivation* ($r = 0.33$) and Q2 - *Motivation to work fast* ($r = 0.44$), the item Q3 - *Feeling of fun* achieves only a weak effect ($r = 0.24$). Thus, hypothesis H2a can be supported, and it is assumed that the perceived motivation of the subjects can be increased by competitive audio-gamification.

Table 5 presents the results of the mean value comparisons for the evaluation of H2b. The Mann-Whitney-U-Test provides statistically significant results for comparing the cooperative group and control group for all three items. In addition, the effect size for the two items Q1 - *General Motivation* ($r = 0.35$) and Q3 - *Feeling of fun* ($r = 0.35$) reach a medium value, and for Q2 - *Motivation to work fast*, a high effect size value ($r = 0.51$). Thus, H2b is also supported, and it can be assumed that cooperative gamification has a positive effect on the perceived motivation.

4.3. Comparison of cooperative vs. competitive

In the descriptive results in Table 2, the competitive group completed the tasks on average faster than the cooperative experimental group. Thus, our hypothesis H3 (subjects in the cooperative group are faster than in the competitive group) cannot be supported as it was set up initially with an opposite expected result. Therefore, we accept the null hypothesis.

Finally, we check our hypothesis H4 (subjects in the cooperative group are more motivated than in the competitive group). The descriptive data for the three items in Table 3 show that the cooperative group achieves a higher mean value for all questionnaire items than the competitive group. The results of the statistical analysis are summarized in Table 6.

Comparing the mean values leads to no statistically significant results for all three items. Thus, H4 cannot be supported, and it must be assumed that it has no statistically significant influence on the motivation whether a cooperative or competitive design is applied.

4.4. Additional statistical analyses

In addition to the hypotheses, it was also examined whether an influence of the factors age, gender, or previous knowledge in dealing with

Table 3

Mean and standard deviation for all three motivational items (General motivation/Motivation to work faster/Feeling of fun) per experimental group.

| Condition | General Motivation | | | Motivation to work faster | | | Feeling of fun | | |
|-------------|--------------------|------|------|---------------------------|------|------|----------------|------|------|
| | M | Mdn | SD | M | Mdn | SD | M | Mdn | SD |
| Control | 2.64 | 3.00 | 1.22 | 2.23 | 2.00 | 1.02 | 3.00 | 3.00 | 1.20 |
| Competitive | 3.38 | 4.00 | 0.98 | 3.38 | 4.00 | 1.30 | 3.58 | 4.00 | 0.95 |
| Cooperative | 3.52 | 4.00 | 1.06 | 3.58 | 4.00 | 1.23 | 3.76 | 4.00 | 1.20 |

Table 4

Results of the Mann-Whitney-U-Test and effect size for the comparison between the control group and the competitive group regarding the three items for measuring motivation.

| Item | U | z | p (one-tailed) | r |
|----------------------------|--------|-------|----------------|-----|
| Q1 General Motivation | 180.50 | -2.29 | .011 | .33 |
| Q2 Motivation to work fast | 142.00 | -3.06 | .001 | .44 |
| Q3 Feeling of fun | 210.50 | -1.66 | .049 | .24 |

Table 5

Results of the Mann-Whitney-U-Test and effect size for the comparison between the control group and the cooperative group concerning the three items for measuring motivation.

| Item | U | z | p (one-tailed) | r |
|----------------------------|--------|-------|----------------|-----|
| Q1 General Motivation | 217.50 | -2.59 | .005 | .35 |
| Q2 Motivation to work fast | 149.50 | -3.76 | .000 | .51 |
| Q3 Feeling of fun | 231.50 | -2.36 | .010 | .35 |

Table 6

Results of the Mann-Whitney-U-Test for the comparison between the cooperative and the competitive group regarding the three items for measuring motivation.

| Item | U | Z | p (one-tailed) |
|----------------------------|-------|-------|----------------|
| Q1 General Motivation | 409.0 | -0.32 | .373 |
| Q2 Motivation to work fast | 395.5 | -0.53 | .298 |
| Q3 Feeling of fun | 363.5 | -1.06 | .146 |

IVAs on the three motivation items and the time required could be determined. The reason is to control for possible influences of extraneous variables. The correlations were calculated for the entire data set. Since the three items used to measure motivation are not normally distributed, as shown in section 4.2., a Spearman correlation is calculated. The analysis shows no significant correlations between age, gender, or the experience with IVAs and one of the other variables. Interestingly, the three questions concerning motivation correlate, suggesting good construct validity. The results are summarized in Table 7. Negative correlation values for gender are in favor of male

Table 7

Correlation matrix of the additional statistical analyses. P-values are two-tailed.

| | Age | Gender | Experience with IVAs | General Motivation | Motivation to work fast | Feeling of fun | Adjusted required time |
|-------------------------|-----|--------|----------------------|--------------------|-------------------------|----------------|------------------------|
| Age | r | 1.00 | -.196 | .034 | -.130 | -.017 | -.211 |
| | p | | .079 | .761 | .248 | .882 | .927 |
| Gender | r | | 1.00 | -.190 | .206 | .211 | -.022 |
| | p | | | .089 | .065 | .059 | .844 |
| Experience with IVAs | r | | | 1.00 | -.109 | .129 | -.027 |
| | p | | | | .331 | .923 | .253 |
| General Motivation | r | | | | 1.00 | .387** | .527** |
| | p | | | | | .000 | .373 |
| Motivation to work fast | r | | | | | 1.00 | .405** |
| | p | | | | | | .000 |
| Feeling of fun | r | | | | | | 1.00 |
| | p | | | | | | |
| Adjusted required time | r | | | | | | |
| | p | | | | | | |

subjects, and positive correlation values are in favor of females.

The variables age and gender appear interesting in the context of the study. Reminder, 53 subjects are male, and 28 are female. When conducting the study, males were on average 24.19 years old ($SD = 4.57$) and females 22.36 years old ($SD = 3.91$). The age distribution between the genders within the three experimental conditions is unremarkable. If we take the results of Table 2 and further analyze the results according to gender, we get the following results, as displayed in Table 8.

Although the results do not become statistically significant with the t-test, strong trends are apparent. In the control and the cooperative conditions, the male subjects took more time to accomplish the tasks than the female subjects. In contrast, female subjects were slower in the competitive condition. A more balanced gender distribution per condition might have led to more conclusive results in this case. Still, this is an exciting result, which shows that gender effects can play an essential role in performance, which also happened to be the case on other occasions (Bräuer & Mazarakis, 2019). Age- and gender-related differences in the use of voice-activated artificially intelligent (voice-AI) devices have not gone unnoticed and are already being addressed in studies (Zellou, Cohn, & Ferenc Segedin, 2021). The need for future research in this area should already be noted at this point.

For the motivational aspects (general motivation and motivation to work fast), it is already evident from the correlation table that the female subjects were more motivated than the males. The questionnaire items for general motivation were tended to be rated lower by males ($M = 3.08$, $SD = 1.14$, $Mdn = 3.00$) than for females ($M = 3.54$, $SD = 1.08$, $Mdn = 4.00$), $U = 564.50$, $z = -1.84$, $p = .065$. In addition, the questionnaire items for the motivation to work fast were also tended to be

Table 8

Number of subjects, as well as the mean, median, and standard deviation of the adjusted required time per experimental group and gender.

| Condition | N | M | Mdn | SD | p (two-tailed) |
|--------------------|----|--------|--------|--------|----------------|
| Control Male | 16 | 654.65 | 646.27 | 100.40 | .122 |
| Control Female | 6 | 584.70 | 587.53 | 51.03 | |
| Competitive Male | 16 | 473.04 | 466.49 | 104.10 | .269 |
| Competitive Female | 10 | 518.10 | 514.82 | 89.12 | |
| Cooperative Male | 21 | 552.77 | 549.62 | 113.86 | .069 |
| Cooperative Female | 12 | 482.20 | 472.42 | 82.08 | |

rated lower by males ($M = 2.94$, $SD = 1.34$, $Mdn = 3.00$) than for females ($M = 3.54$, $SD = 1.20$, $Mdn = 4.00$), $U = 557.00$, $z = -1.89$, $p = .060$. Again, a more balanced gender distribution may have resulted in more definitive answers in this context. However, between the three experimental conditions, the results are unexceptional.

5. Discussion

This pioneering study was designed to investigate a new innovative field of application of gamification and IVAs for human behavior, which we call audio-gamification. The goal was to systematically determine whether game design elements also have a motivating and performance-enhancing effect in purely audio-based applications for IVAs. Following our expectations, a statistically significant effect on the motivation of the subjects as well as on the speed at which they complete given tasks could be determined. Thus the results are in accordance with previous gamification research in other fields (Johnson et al., 2016; Majuri et al., 2018; Seaborn & Fels, 2015; Warmelink et al., 2018).

As Murad, Munteanu, Clark, and Cowan (2018) point out, feedback is of central importance in the design of IVAs. Audio-gamification can help to provide the user with the necessary feedback in a familiar way. At the same time, care must be taken not to overwhelm the user with additional elements which could distract or bore them. Our study is compliant with design recommendations for non-complex interaction models (Murad et al., 2018), and our results show that audio-gamification can improve the experience with IVAs. Our results provide implications for improving user motivation of IVAs, which might be helpful in many application domains. Possible areas of application could be, for example, language learning, which is currently a popular field of research on the use of IVAs (Dizon, 2021). Since it is known that gamification can improve both motivation for learning and learning success (Bai et al., 2020; Majuri et al., 2018; Sailer & Homner, 2020), audio-gamification could possibly also produce such an effect in learning with IVAs. Various applications for audio-gamification are also possible in the realm of sustainability research. Drivers could be incentivized to adopt ecologically friendly driving behavior, for example, by incorporating game design features in cars, similar to the Landesberger et al. (2020) study. Alternatively, in the application in the smart home, users could be motivated to use the devices in the household in a more energy-efficient way (Mendez et al., 2020).

Research into the effects of audio-gamification opens new possibilities for the further development of IVAs. Another field of application in which gamification can be implemented without visual support is applications for visually impaired individuals (Mazarakis, 2021; Sciarretta & Alimenti, 2021). Through an audio-only implementation, obstacles such as the use of color to convey instruction, to indicate an action, or to prompt a response, which often occur during the implementation of gamification, can be bypassed (Smith & Abrams, 2019). In this respect, our results help to go beyond the work of Tejedor-Garcia et al. (2020) to show that for the implementation of gamification in the context of IVAs, a smartphone or other device with a display is not necessarily required. To enable the most inclusive design of IVAs, we create new possibilities to make applications more appealing to people with limited vision by using only sound and speech to represent the game design elements. Our study provides valuable insight that offers a new perspective on the potential of gamification in a wide range of applications.

The positive results of our study can also be attributed to the successful embedding of the tasks in the narrative around the AI TILA. With the help of the narrative, the user was guided through the menu of the application. As a result, receiving and confirming the tasks was more like a conversation with TILA. This also is in line with the results of Ji et al. (2017), according to which users prefer natural conversations over menu-based navigation. This leads to the assumption that by including a narrative element, an enormous potential can be unlocked regarding the improvement of the acceptance of IVAs.

5.1. Limitations

In addition to the effect of audio-gamification on motivation and performance, this study examined whether differences between cooperative and competitive game design elements could be demonstrated. Contrary to the results of, e.g., Morschheuser et al. (2019) and our expectations, no difference in motivation or performance could be detected between the groups. Other studies also struggled with the differences between cooperative and competitive game design elements, e.g., they were not always clear either (Dindar et al., 2021; Marinho et al., 2019). In the case of our study, there may be too marginal differences in the design of the two groups. The same five game design elements (points, time pressure, level, ranking, and narrative) were used in both groups. The only difference in terms of game design elements and besides the different amount of information provided between the cooperative and the competitive group was the implementation of the ranking. To achieve a clear impact, the differences between cooperative and competitive gamification should be highlighted more clearly. For example, the narrative could be adapted more or challenges designed cooperatively or competitively could be used.

A possible general limitation should not remain unmentioned: a cooperative setting may not always be beneficial. Despite the efforts of, e.g., Morschheuser (2019), others provide different findings. A highly competitive game design element like a leaderboard or ranking can have a massive impact on performance (Bräuer & Mazarakis, 2019). In addition, Dindar et al. (2021) show that subjects in their study completed the tasks in the competitive condition statistically significant faster than in the cooperative condition. Therefore, the context in which gamification is to be used must always be considered (Bräuer & Mazarakis, 2019; Finckenhagen, 2017; Koivisto & Hamari, 2014; Mazarakis, 2021; Richards, Thompson, & Graham, 2014).

In our study, the motivation of the subjects was determined based on a questionnaire with three self-developed questions. The procedure for measuring motivation could have been based on established questionnaire tools, such as the Intrinsic Motivation Inventory (Intrinsic Motivation Inventory (IMI) – Selfdeterminationtheory.Org, n.d.). The advantage of using such a tool would be, besides the proven validity of the questionnaire, the possibility to collect information about the satisfaction of different basic needs (autonomy, relatedness, and competence). This would allow more precise conclusions to be drawn about the effect of audio-gamification on motivation.

Also, the effect that the individual game design elements might have on each other was not considered in the study. As Mekler et al. (2017), and Mazarakis and Bräuer (2018, 2022) were able to show, different game design elements have quite different effects when applied individually. Since the focus of our study was to test whether an effect can be achieved at all, several game design elements were used in combination so that no conclusions could be drawn about the mode of action of the individual game design elements. Especially about the audio feedback, information on the individual effects could be extremely helpful for the design of skills. In addition, it may be necessary to take individual needs into account when designing audio game design elements and to offer the user the possibility to adjust or switch off the elements.

As with most studies in the field of gamification, there is a lack of evaluation of the long-term impact of gamification (Mazarakis, 2021; Rapp et al., 2019; Seaborn & Fels, 2015). The effect investigated was demonstrated in a laboratory setting in which the subjects interacted with the system once and for no longer than half an hour. Therefore, it remains questionable how audio-gamification would affect motivation if a user used a similar incentive system several times over a longer period of time. To answer this question, a follow-up study should be conducted. In such a study, however, the setting would have to be adapted since the skill developed for the laboratory experiment with the predefined tasks would not be reusable in a long-term field study. Alternatively, the audio-gamification concept might be transferred to a foreign language learning application, for example. Another imaginable scenario to

implement the research design in a less restrictive laboratory setting would be to emphasize health exercises, specifically relaxation exercises, emphasizing breathing exercises.

A final limiting aspect in conducting the study is the relatively low proportion of daily users of IVAs (16.67%) in the sample. Although the ratio of subjects without experience with IVAs compared to subjects with experience in our study corresponds to the prevalence of experience with IVAs in Germany (Kinsella, 2021), it is still possible that the usage behavior could be influenced by this aspect. In addition, while no major differences in interaction with the skill between experienced and inexperienced users were found (see Table 7), future studies should include a larger group of users with more prior experience, e.g., daily use. This could lead to group-specific insight, e.g., different results for power users.

5.2. Future research

Our study provides first results in the field of audio-gamification and shows the potential of integrating game design elements without the use of a screen. At the same time, this study raises several further questions that should be addressed by future research to provide a valid basis for the design of IVAs based on our results. A first starting point is provided by the additionally considered differences between the genders. In both fields of gamification and research on IVAs, initial studies address the effects of age and gender (Codish & Ravid, 2017; Jent & Janneck, 2018; Zellou et al., 2021). Our data also suggest that there might be differences in the motivational effect of audio-gamification due to these influencing factors. Similarly, users' prior experience with IVAs could also be an essential factor influencing the impact of audio gamification. Future research could address this potentially limiting factor.

Many users of Amazon Alexa are annoyed after some time because the IVA always gives them the same long answers. Amazon has therefore now built in the option to receive only short answers such as "yes" or a tone instead of a spoken answer (Alexa Brief Mode Explained, 2019). Since the addition of game design elements automatically extends the responses of the IVA, it is possible that a similar fatigue effect can result. By conducting an online study over a longer period of time, conclusions can be drawn as to whether this effect also occurs with audio game design elements and, if so, what design options there are to counteract this phenomenon.

An important point in designing applications for IVAs is to make the menu navigation as efficient and straightforward as possible (Murad et al., 2018). Unfortunately, the integration of game design elements counteracts this principle and could therefore harm usability. However, since the results on motivation and performance improvement of our study were positive, it should be determined whether the integration of game design elements really has a negative impact on usability. To find out whether such an effect can be verified, this should be investigated using suitable evaluation procedures, such as appropriate questionnaires like the system usability scale (Ghosh, Foong, Zhang, & Zhao, 2018).

A third important point that should be considered for future research is the direct comparison between visual and audio implementation of game design elements. To be able to conclude how important a visual component is in the implementation of game design elements, more studies with different devices and settings are needed. For example, some speakers with IVAs like the Alexa Echo Show or the Google Nest Hub already have integrated a screen. A comparison of the usage behavior between two skills that differ only in the presentation of the game design elements (visual vs. audio) could provide insights into how much users are influenced by a visual component (Montalvo et al., 2021).

One factor that could decisively influence the effect of audio-gamification is the presence of other people in the room. In contrast to working on screens, where users usually use one device alone, several users can simultaneously interact with an IVA (Trajkova &

Martin-Hammond, 2020). The assistants can be trained on different voices and thus identify various users (Amazon.Com Help: About Alexa Voice Profiles, 2020). It has already been shown that working together in a group with IVAs has a positive impact on cooperation within the group and improves the output of group work (Winkler, Söllner, Neuweiler, Conti Rossini, & Leimeister, 2019). On the other hand, the presence of other people can also act as a deterrent and discourage the use of the IVA (Trajkova & Martin-Hammond, 2020). Depending on the implementation of gamification and the personality traits of the user, the presence of other people could lead to amplification or attenuation of the motivating effect.

6. Conclusion

Our pioneering study with 81 subjects empirically investigated the effect of audio-only game design elements on motivation and performance when using an IVA. We answered our research questions regarding how audio-gamification affects motivation and performance. The positive results are consistent with previous studies on the general effect of gamification (Johnson et al., 2016; Majuri et al., 2018; Seaborn & Fels, 2015; Warmelink et al., 2018) and show that a visual component is not necessarily required for the successful implementation of game design elements. We also considered differences in the design of the audio game design elements. However, the comparison between a cooperative and a competitive group did not show any statistically significant difference for our corresponding research questions about motivation and performance. The results thus contradict previous work (Morschheuser et al., 2019) and offer a variety of possibilities to research the contradictions. We believe that our work will contribute to a fundamental paradigm shift for IVAs in developing human-centered skills and creating new, more appealing motivational design approaches, thus possibly promoting the acceptance of IVAs in the long term, especially with gamification.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Credit author statement

Paula Bräuer: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision.; **Athanasios Mazarakis:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

References

- Alexa Brief Mode Explained: How to Turn It on and How It Works. (2019, May 6). The Ambient. <https://www.the-ambient.com/how-to/how-to-use-alexa-brief-mode-500>.
- Allam, A., Kostova, Z., Nakamoto, K., & Schulz, P. J. (2015). The effect of social support features and gamification on a web-based intervention for rheumatoid arthritis patients: Randomized controlled trial. *Journal of Medical Internet Research*, 17(1), e14. <https://doi.org/10.2196/jmir.3510>
- Amazon.Com help: About Alexa voice Profiles. (2020). <https://www.amazon.com/gp/help/customer/display.html?nodeId=202199440>.
- Antin, J., & Churchill, E. F. (2011). Badges in Social Media: A Social Psychological Perspective. *CHI 2011 Gamification Workshop Proceedings*, 1–4.
- Atkinson, T., Baier, H., Coppleson, T., Devlin, S., & Swan, J. (2019). The text-based adventure AI competition. *IEEE Transactions on Games*, 11(3), 260–266. <https://doi.org/10.1109/TG.2019.2896017>
- Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and Synthesis of qualitative data in educational contexts. *Educational Research Review*, 30, 100322. <https://doi.org/10.1016/j.edurev.2020.100322>
- Baneres, D., Rodriguez, M. E., & Serra, M. (2019). An early feedback prediction system for learners at-risk within a first-year higher education course. *IEEE Transactions on Learning Technologies*, 12(2), 249–263. <https://doi.org/10.1109/TLT.2019.2912167>

- de Barcelos Silva, A., Gomes, M. M., da Costa, C. A., da Rosa Righi, R., Barbosa, J. L. V., Pessin, G., et al. (2020). Intelligent personal assistants: A systematic literature review. *Expert Systems with Applications*, 147, 113193. <https://doi.org/10.1016/j.eswa.2020.113193>
- Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: Changing user needs and motivations. *Interactions*, 25(5), 38–43. <https://doi.org/10.1145/3236669>
- Bräuer, P., & Mazarakis, A. (2019). Badges or a leaderboard? How to gamify an augmented reality warehouse setting. In J. Koivisto, & Hamari (Eds.), *Vol. 2359. Proceedings of the 3rd international GamiFIN conference—GamiFIN 2019* (pp. 229–240) (CEUR-WS.org).
- Cechetti, N. P., Bellei, E. A., Biduski, D., Rodriguez, J. P. M., Roman, M. K., & De Marchi, A. C. B. (2019). Developing and implementing a gamification method to improve user engagement: A case study with an M-health application for hypertension monitoring. *Telematics and Informatics*, 41, 126–138. <https://doi.org/10.1016/j.tele.2019.04.007>
- Clark, L. (2003). Speaking pictures: The role of sound and orality in audio presentations of children's picture books. *New Review of Children's Literature and Librarianship*, 9(1), 1–19. <https://doi.org/10.1080/1361454032000232077>
- Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., et al. (2019). The state of speech in HCI: Trends, themes and challenges. *Interacting with Computers*, 31(4), 349–371. <https://doi.org/10.1093/iwc/iwz016>
- Codish, D., & Ravid, G. (2017). *Gender Moderation in Gamification: Does One Size Fit All?* Hawaii International Conference on System Sciences. <https://doi.org/10.24251/HICSS.2017.244>
- Collins, K. (2008). *From Pac-Man to pop music: Interactive audio in games and new media*. Ashgate.
- Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., et al. (2017). "What can I help you with?": Infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1–12). <https://doi.org/10.1145/3098279.3098539>
- Csikszentmihalyi, M. (2009). *Flow: The psychology of optimal experience*. Harper Collins.
- Czeszumski, A., Ehinger, B. V., Wahn, B., & König, P. (2019). The social situation affects how we process feedback about our actions. *Frontiers in Psychology*, 10(361). <https://doi.org/10.3389/fpsyg.2019.00361>
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Dellaert, B. G. C., Shu, S. B., Arentze, T. A., Baker, T., Diehl, K., Donkers, B., et al. (2020). Consumer decisions with artificially intelligent voice assistants. *Marketing Letters*. <https://doi.org/10.1007/s11002-020-09537-5>
- Derby, J. L., Rickel, E. A., Harris, K. J., Lovell, J. A., & Chaparro, B. S. (2020). "We didn't catch that!" using voice text input on a mixed reality headset in noisy environments. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting*, 64(1), 2102–2106. <https://doi.org/10.1177/107181320641509>
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). *From Game Design Elements to Gamefulness: Defining "Gamification."* Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, 9–15. <https://doi.org/10.1145/2181037.2181040>
- Dindar, M., Ren, L., & Järvenoja, H. (2021). An experimental study on the effects of gamified cooperation and competition on English vocabulary learning. *British Journal of Educational Technology*, 52(1), 142–159. <https://doi.org/10.1111/bjet.12977>
- Dizon, G. (2021). Affordances and constraints of intelligent personal assistants for second-language learning. *RELC Journal*. <https://doi.org/10.1177/00336882211020548>, 003368822110205.
- Dunn, J. (2016, August 26). Virtual assistants like Siri and Alexa look poised to explode. *Business Insider Deutschland*. <https://www.businessinsider.de/virtual-assistants-siri-alex-growth-chart-2016-8>.
- Fadhil, A., & Villafiorita, A. (2017). An adaptive learning with gamification & conversational UIs: The rise of CiboPoliBot. *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, 408–412. <https://doi.org/10.1145/3099023.3099112>
- Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- Finckenhagen, K. R. (2017). *Context in Gamification—Contextual Factors and Successful Gamification*. https://www.ntnu.edu/documents/139799/1279149990/09+Article+Final_karolirf_fors%C3%B8k_2017-12-08-05-35-05_TPD4505.Karoline.Fincke_nhagen.pdf/4f269657-0e54-4999-9de8-6ba8e109d386.
- Friberg, J., & Gårdenfors, D. (2004). Audio games: New perspectives on game audio. *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology - ACE '04*, 148–154. <https://doi.org/10.1145/1067343.1067361>
- Garcia, F. E., & de Almeida Neris, V. P. (2013). Design guidelines for audio games. In M. Kurosu (Ed.), *8005. Proceedings of the 15th international conference on human-computer interaction* (pp. 229–238). Springer. https://doi.org/10.1007/978-3-642-39262-7_26.
- Geelan, B., De Salas, K., Lewis, I., King, C., Edwards, D., & O'Mara, A. (2015). Improving learning experiences through gamification: A case study. *Australian Educational Computing*, 30(1), 1–21.
- Ghosh, D., Foong, P. S., Zhang, S., & Zhao, S. (2018). Assessing the utility of the system usability scale for evaluating voice-based user interfaces. In *Proceedings of the Sixth International Symposium of Chinese CHI on - ChineseCHI '18* (pp. 11–15). <https://doi.org/10.1145/3202667.3204844>
- Grobely, J., Smierzchalska, J., & Czapkowski, K. (2018). Narrative gamification as a method of increasing sales performance: A field experimental study. *International Journal of Academic Research in Business and Social Sciences*, 8(3), 430–447. <https://doi.org/10.6007/IJARBS/v8-i3/3940>
- Groening, C., & Binnewies, C. (2019). "Achievement unlocked!"—the impact of digital achievements as a gamification element on motivation and performance. *Computers in Human Behavior*, 97, 151–166. <https://doi.org/10.1016/j.chb.2019.02.026>
- Guardia, L., J. G., Ryan, R. M., Couchman, C. E., & Deci, E. L. (2000). Within-person variation in security of attachment: A self-determination theory perspective on attachment, need fulfillment, and well-being. *Journal of Personality and Social Psychology*, 79(3), 367–384. <https://doi.org/10.1037/0022-3514.79.3.367>
- Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, 90, 343–350. <https://doi.org/10.1016/j.chb.2018.08.009>
- Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *47th Hawaii International Conference on System Sciences (HICSS)* (pp. 3025–3034). <https://doi.org/10.1109/HICSS.2014.377>
- Harris, R. A. (2004). *Voice interaction design: Crafting the new conversational speech systems*. Elsevier.
- Hofmann, H., Tobisch, V., Ehrlich, U., Berton, A., & Mahr, A. (2014). Comparison of speech-based in-car hmi concepts in a driving simulation study. In *Proceedings of the 19th International Conference on Intelligent User Interfaces* (pp. 215–224). <https://doi.org/10.1145/2557500.2557509>
- Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly*, 37(1), 81–88. <https://doi.org/10.1080/02763869.2018.1404391>
- Intrinsic Motivation Inventory (IMI) – selfdeterminationtheory.org. (n.d.). Retrieved November 6, 2020, from <https://selfdeterminationtheory.org/intrinsic-motivation-inventory/>.
- Jent, S., & Janneck, M. (2018). Using gamification to enhance user motivation: The influence of gender and age. In L. E. Freund, & W. Cellary (Eds.), *Vol. 601. Advances in the human side of service engineering* (pp. 3–10). Springer International Publishing. https://doi.org/10.1007/978-3-319-60486-2_1
- Ji, J., Wang, Q., Battad, Z., Gou, J., Zhou, J., Divekar, R., et al. (2017). A two-layer dialogue framework for authoring social bots. *Alexa Prize Proceedings*.
- 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. <https://doi.org/10.1016/j.invent.2016.10.002>
- Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. John Wiley & Sons.
- Keusch, F., & Zhang, C. (2017). A review of issues in gamified surveys. *Social Science Computer Review*, 35(2), 147–166. <https://doi.org/10.1177/0894439315608451>
- Kinsella, B. (2019, March 7). U.S. Smart display user base grew 558% in 2018 and more than doubled in second half of the year, amazon holds two-thirds market share. *Voicebot.Ai*. <https://voicebot.ai/2019/03/07/u-s-smart-display-user-base-grew-558-in-2018-and-more-than-doubled-in-second-half-of-the-year-amazon-holds-two-thirds-market-share/>.
- Kinsella, B. (2021, June 17). Germany smart speaker adoption closely mirrors U.S. Pattern—new report with 30+ charts. *Voicebot.Ai*. <https://voicebot.ai/2021/06/17/germany-smart-speaker-adoption-closely-mirrors-u-s-pattern-new-report-with-30-charts/>.
- Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior*, 35, 179–188. <https://doi.org/10.1016/j.chb.2014.03.007>
- Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, 45, 191–210. <https://doi.org/10.1016/j.ijinfomgt.2018.10.013>
- 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. <https://doi.org/10.1177/1046878118774385>
- Landers, R. N., Bauer, K. N., & Callan, R. C. (2015). Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior*, 71, 508–515. <https://doi.org/10.1016/j.chb.2015.08.008>
- Landers, R. N., Collmus, A. B., & Williams, H. (2019). The greatest battle is within ourselves: An experiment on the effects of competition alone on task performance. *International Journal of Human-Computer Studies*, 127, 51–61. <https://doi.org/10.1016/j.ijhcs.2018.09.011>
- Landesberger, J., Ehrlich, U., & Minker, W. (2020). "What is it?" How to Collect Urgent Utterances using a Gamification Approach. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 19–22). <https://doi.org/10.1145/3409251.3411713>
- Leaning, M. (2015). A study of the use of games and gamification to enhance student engagement, experience and achievement on a theory-based course of an undergraduate media degree. *Journal of Media Practice*, 16(2), 155–170. <https://doi.org/10.1080/14682753.2015.1041807>
- Li, C., & Yanagisawa, H. (2021). Intrinsic motivation in virtual assistant interaction for fostering spontaneous interactions. *PLoS One*, 16(4). <https://doi.org/10.1371/journal.pone.0250326>. e0250326.
- Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., et al. (2019). Talk to me: Exploring user interactions with the amazon Alexa. *Journal of Librarianship and Information Science*, 51(4), 984–997. <https://doi.org/10.1177/0961000618759414>
- Lopatovska, I., & Williams, H. (2018). Personification of the Amazon Alexa: BFF or a Mindless Companion. In *18. Proceedings of the 2018 Conference on Human Information Interaction & Retrieval - CHIIR* (pp. 265–268). <https://doi.org/10.1145/3176349.3176868>

- Majuri, J., Koivisto, J., & Hamari, J. (2018). Gamification of education and learning: A review of empirical literature. In , 2018. *Proceedings of the 2nd International GamiFIN Conference, GamiFIN* (pp. 11–19).
- Marinho, A., Oliveira, W., Bittencourt, I. I., & Demerval, D. (2019). Does Gamification Improve Flow Experience in Classroom? An Analysis of Gamer Types in Collaborative and Competitive Settings. *Revista Brasileira de Informática Na Educação*, 27(2), 40. <https://doi.org/10.5753/rbie.2019.27.02.40>
- Mazarakis, A. (2015). Using gamification for technology enhanced learning: The case of feedback mechanisms. *Bulletin of the IEEE Technical Committee on Learning Technology*, 4(17), 6–9.
- Mazarakis, A. (2021). Gamification reloaded: Current and future trends in gamification science. *I-com*, 20(3), 279–294. <https://doi.org/10.1515/icom-2021-0025>
- Mazarakis, A., & Bräuer, P. (2018). Gamification is working, but which one exactly? Results from an experiment with four game design elements. *Proceedings of the Technology, Mind, and Society*, 22(1). <https://doi.org/10.1145/3183654.3183667>
- Mazarakis, A., & Bräuer, P. (2020). Gamification of an open access quiz with badges and progress bars: An experimental study with scientists. In J. Koivisto, M. Bujic, & J. Hamari (Eds.), Vol. 2637. *Proceedings of the 4th international GamiFIN conference—GamiFIN 2020* (pp. 62–71) (CEUR-WS.org).
- Mazarakis, A., & Bräuer, P. (2022). Gamification is working, but which one exactly? Results from an experiment with four game design elements. *International Journal of Human-Computer Interaction*, 1–16. <https://doi.org/10.1080/10447318.2022.2041909>
- McGonigal, J. (2011). *Reality is broken*. Penguin Press.
- Mekler, E. D., Bruehlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, 71, 525–534. <https://doi.org/10.1016/j.chb.2015.08.048>
- Mendez, J. I., Ponce, P., Mata, O., Meier, A., Peffer, T., Molina, A., et al. (2020). Empower saving energy into smart homes using a gamification structure by social products. In 2020 *IEEE International Conference on Consumer Electronics (ICCE)* (pp. 1–7). <https://doi.org/10.1109/ICCE46568.2020.9043174>
- Mewes, A., Hensen, B., Wacker, F., & Hansen, C. (2017). Touchless interaction with software in interventional radiology and surgery: A systematic literature review. *International Journal of Computer Assisted Radiology and Surgery*, 12(2), 291–305. <https://doi.org/10.1007/s11548-016-1480-6>
- Montalvo, F., Schreck, J., Krause, K., Sasser, J., McConnell, D. S., & Smither, J. A. (2021). Do touchscreens make intelligent personal assistants more useable? An initial usability study. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting*, 65(1), 899–903. <https://doi.org/10.1177/1071181321651283>
- 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. <https://doi.org/10.1016/j.ijhcs.2018.10.001>
- Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), Vol. 2. *Handbook of research on educational communications and technology* (pp. 745–783). Lawrence Erlbaum Associates.
- Murad, C., Munteanu, C., Clark, L., & Cowan, B. R. (2018). Design guidelines for hands-free speech interaction. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '18* (pp. 269–276). <https://doi.org/10.1145/3236112.3236149>
- Newman, N. (2018). The Future of Voice and the Implications for News. <https://reutersinstitute.politics.ox.ac.uk/our-research/future-voice-and-implications-news>.
- NPR and Edison Research. (2020). *Smart Audio Report*. <https://www.nationalpublicmedia.com/uploads/2020/04/The-Smart-Audio-Report-Spring-2020.pdf>.
- Nykanen, A., Lopez, M., & Toulson, R. (2016). Safe and sound drive: Design of interactive sounds supporting energy efficient behaviour. *Interactive Audio Systems Symposium, York*, 23 September 2016.
- Pe-Than, E. P. P., Goh, D. H.-L., & Lee, C. S. (2014). Making work fun: Investigating antecedents of perceived enjoyment in human computation games for information sharing. *Computers in Human Behavior*, 39, 88–99. <https://doi.org/10.1016/j.chb.2014.06.023>
- Portal. (2019). In Wikipedia. [https://de.wikipedia.org/w/index.php?title=Portal_\(Computerspiel\)&oldid=185754590](https://de.wikipedia.org/w/index.php?title=Portal_(Computerspiel)&oldid=185754590).
- 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. <https://doi.org/10.1037/a0019440>
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4–13. <https://doi.org/10.1002/bs.3830280103>
- Rapp, A., Hopfgartner, F., Hamari, J., Linehan, C., & Cena, F. (2019). Strengthening gamification studies: Current trends and future opportunities of gamification research. *International Journal of Human-Computer Studies*, 127, 1–6. <https://doi.org/10.1016/j.ijhcs.2018.11.007>
- Richards, C., Thompson, C. W., & Graham, N. (2014). Beyond designing for motivation: The importance of context in gamification. *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '14)*, 217–226. <https://doi.org/10.1145/2658537.2658683>
- Röber, N., & Masuch, M. (2005). *Leaving the Screen New Perspectives in Audio-Only Gaming*. <https://smartech.gatech.edu/handle/1853/50168>.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54–67. <https://doi.org/10.1006/ceps.1999.1020>
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.
- 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. <https://doi.org/10.1007/s11031-006-9051-8>
- Sailer, M., Hense, J., Mandl, H., & Klevers, M. (2017). Fostering development of work competencies and motivation via gamification. In M. Mulder (Ed.), Vol. 23. *Competence-based vocational and professional education* (pp. 795–818). Springer. https://doi.org/10.1007/978-3-319-41713-4_37
- 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. <https://doi.org/10.1016/j.chb.2016.12.033>
- Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32(1), 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Schmidmaier, M., Hußmann, H., & Runge, D. M. (2020). Beep beep: Building trust with sound. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–8). <https://doi.org/10.1145/3334480.3382848>
- Sciarretta, E., & Alimenti, L. (2021). Smart speakers for inclusion: How can intelligent virtual assistants really assist everybody? In M. Kurosu (Ed.), Vol. 12762. *Human-computer interaction. Theory, methods and tools* (pp. 77–93). Springer International Publishing. https://doi.org/10.1007/978-3-030-78462-1_6
- Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74, 14–31. <https://doi.org/10.1016/j.ijhcs.2014.09.006>
- Shang, J., & Wu, J. (2019). Enabling secure voice input on augmented reality headsets using internal body voice. In 2019 *16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)* (pp. 1–9). <https://doi.org/10.1109/SAHCN.2019.8824980>
- Shepherd, C., & Liu, N. (2019, October 4). Baidu bets big on smart speakers. *Financial Times*, 17.
- Silva-Coira, F., Cortiñas, A., & Pedreira, O. (2016). Intelligent virtual assistant for gamified environments. *PACIS 2016 Proceedings*.
- Smith, K., & Abrams, S. S. (2019). Gamification and accessibility. *International Journal of Information and Learning Technology*, 36(2), 104–123. <https://doi.org/10.1108/IJILT-06-2018-0061>
- Speech Synthesis Markup Language (SSML) Reference | Alexa Skills Kit. (n.d.). Amazon (Alexa). Retrieved July 17, 2021, from <https://developer.amazon.com/en-US/docs/alexa/custom-skills/speech-synthesis-markup-language-ssml-reference.html>.
- Stanculescu, L. C., Bozzon, A., Sips, R.-J., & Houben, G.-J. (2016). Work and Play: An Experiment in Enterprise Gamification. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)* (pp. 346–358). <https://doi.org/10.1145/2818048.2820061>
- Steinhaeuser, S. C., Schaper, P., & Lugin, B. (2021). Comparing a robotic storyteller versus audio book with integration of sound effects and background music. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 328–333). <https://doi.org/10.1145/3434074.3447186>
- Tejedor-Garcia, C., Escudero-Mancebo, D., Cardenoso-Payo, V., & Gonzalez-Ferreras, C. (2020). Using challenges to enhance a learning game for pronunciation training of English as a second language. *IEEE Access*, 8, 74250–74266. <https://doi.org/10.1109/ACCESS.2020.2988406>
- Tejedor-Garcia, C., Escudero-Mancebo, D., Gonzalez-Ferreras, C., Cámara-Arenas, E., & Cardenoso-Payo, V. (2016). Improving L2 production with a gamified computer-assisted pronunciation training tool, *tiptoptalk! Proc. IberSPEECH*, 177–186.
- Thiebes, S., Lins, S., & Basten, D. (2014). Gamifying information systems—a Synthesis of gamification mechanics and dynamics. *ECIS 2014 Proceedings*, 1–17.
- Trajkova, M., & Martin-Hammond, A. (2020). “Alexa is a toy”: Exploring older adults’ reasons for using, limiting, and abandoning Echo. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). <https://doi.org/10.1145/3313831.3376760>
- Vansteenkiste, M., & Ryan, R. M. (2013). On psychological growth and vulnerability: Basic psychological need satisfaction and need frustration as a unifying principle. *Journal of Psychotherapy Integration*, 23(3), 263–280. <https://doi.org/10.1037/a0032359>
- Vtyurina, A., & Fournay, A. (2018). Exploring the role of conversational cues in guided task support with virtual assistants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–7). <https://doi.org/10.1145/3173574.3173782>
- Warmelink, H., Koivisto, J., Mayer, I., Vesa, M., & Hamari, J. (2018). Gamification of the work floor: A literature review of gamifying production and logistics operations. In *Proceedings of the 51st Hawaii International Conference on System Sciences* (pp. 1108–1117).
- Warmelink, H., Koivisto, J., Mayer, I., Vesa, M., & Hamari, J. (2020). Gamification of production and logistics operations: Status quo and future directions. *Journal of Business Research*, 106, 331–340. <https://doi.org/10.1016/j.jbusres.2018.09.011>
- Weber, P., & Ludwig, T. (2020). (Non-)Interacting with conversational agents: Perceptions and motivations of using chatbots and voice assistants. In *Proceedings of the Conference on Mensch Und Computer* (pp. 321–331). <https://doi.org/10.1145/3404983.3405513>
- Winkler, R., Söllner, M., Neuweiler, M. L., Conti Rossini, F., & Leimeister, J. M. (2019). Alexa, can you help us solve this problem?: How conversations with smart personal assistant tutors increase task group outcomes. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19* (pp. 1–6). <https://doi.org/10.1145/3290607.3313090>

Zellou, G., Cohn, M., & Ferenc Segedin, B. (2021). Age- and gender-related differences in speech alignment toward humans and voice-AI. *Frontiers in Communication*, 5, 600361. <https://doi.org/10.3389/fcomm.2020.600361>

Zhang, N., Mi, X., Feng, X., Wang, X., Tian, Y., & Qian, F. (2019). Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions

on virtual personal assistant systems. In *2019 IEEE Symposium on Security and Privacy (SP)* (pp. 1381–1396). <https://doi.org/10.1109/SP.2019.00016>

Zichermann, G., & Cunningham, C. (2011). *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. O'Reilly Media.

CHAPTER

FIVE

Audio-Gamification Field Experiment

The study presented in this chapter, like the other two studies, was also published in a peer-reviewed journal:

Bräuer, P., & Mazarakis, A. (2022). How to Design Audio-Gamification for Language Learning with Amazon Alexa? - A Long-Term Field Experiment. *International Journal of Human-Computer Interaction*.
<https://doi.org/10.1080/10447318.2022.2160228>

Impact Factor 2022: 4.7

The conception of the study was done jointly by the author and Dr. Athanasios Mazarakis. The Alexa skill was implemented by Eduard R. as part of his master's thesis together with the author. The data was collected by the author over a period of one year using the Amazon Alexa Platform. The analysis of the data and the writing of the manuscript were done by the author and revised by Dr. Athanasios Marazarkis.

5.1. Preface to Study 3

The effect of individual game design elements on motivation was discussed in Chapter 3, and Chapter 4 showed that it is possible to motivate users of an IVA through audio-gamification alone, even without a screen. Based on these findings, however, the question arises whether the effect of audio-gamification can be reproduced in practice and whether there are differences in the effect depending on the game design elements used.

Previous studies in the areas of gamification (Landers et al., 2018) and IVAs (McTear et al., 2016) have mostly been conducted in laboratory settings. However, this does not

allow us to assess beyond doubt whether the corresponding behavior can be reproduced in real-life situations. In order to obtain the most robust results possible, the study from chapter 4 was complemented by a field study. Such field studies are characterized by the heterogeneity of the users (McTear et al., 2016). A challenge in designing a field study is to make the setting as nonartificial as possible. The scenario in the previous study was explicitly tailored to the circumstances of the experiment, meaning that the tasks were not easily transferable to a real-world setting. Tasks such as vacuuming also occur in everyday life, but it is impossible to use the same metrics to evaluate the tasks because, for example, the size of the home varies greatly, and users would need very different amounts of time to complete the task. For the field study, a new scenario was developed to collect data from users who voluntarily use an Alexa skill in their daily lives. The skill should also represent a non-game context that can be made more motivating through gamification. These requirements could be met by an Alexa vocabulary learning skill. As in the previous lab experiment, and unlike many other studies in the field of voice user interaction research, for this study, we did not use existing applications for IVAs but developed our own Amazon Alexa skill. This Alexa skill was uploaded to all English-language Amazon stores in order to reach the broadest possible audience. What is also unique about this study is that the data collection took place over an entire year. In the field of gamification research in particular, there are repeated calls for long-term studies to test whether the motivational effects of game design elements are also effective in the long term (Cermak-Sassenrath, 2019; Nacke & Deterding, 2017; Thomas et al., 2022). Thus, this study represents a significant contribution not only to this thesis but also to research in the area of gamification and IVAs.

5.2. Technical Aspects of Study 3

In addition to the results described in the publication, this section describes the technical implementation of the Alexa skill for the field study. The Alexa skill was designed to collect data that would reflect real-world user behavior in accordance with Research Question 3. For this purpose, the Alexa skill was made publicly available and uploaded to the Amazon Alexa Store.

The JOVO Framework⁷ (version 3) was used to implement the Alexa skill. The backend containing the skill logic was hosted on AWS Lambda. The collected data was logged at runtime in a MySQL database, also hosted on AWS.

To ensure that users use the skill for as long as possible, an extensive learning set of 500 vocabulary words has been provided. Vocabulary learning requires output in two languages. First, the default system language in which the user interacts with the IVA, and second, the language to be learned. To enable output in languages other than the one the user has selected in the system, the Alexa Skill Kit provides Speech Synthesis Markup Language (SSML)⁸. SSML can be used to add additional information to the text that is to be spoken. For the Alexa skill in this study, SSML was used to recite the vocabulary to the user in German. A code example is shown below:

```
'<speak>
  <voice name="Vicki">
    <lang xml:lang="de-DE">'
      + Hier steht eine deutsche Vokabel +
    '</lang>
  </voice>
</speak>'
```

An intent was created to map the function that checks whether an answer given by the user is correct. The intent was assigned an slot type containing 500 slots representing the correct answers. For slots that are not unique, such as verbs where the user's answer could be "drive" or "to drive," additional synonyms for the slots were created. An explanation of the programming implementation of intents and slots can be found in section 2.2.3.

Unlike the skill developed for Study 2, this Alexa skill can be used by multiple users at the same time. In addition, data is stored so that the Alexa skill can restore the state of the last session when it is restarted, i.e., it is possible to save vocabulary learning progress beyond a single session. Since the logic of the skill is hosted as a web service, care must be taken to keep the data persistent across different sessions. For this purpose, the JOVO framework provides the \$user and \$session properties, which can be used to assign information to a user or session uniquely. A more detailed explanation of this can also be found in section 2.2.3.


⁷ <https://www.jovo.tech/>

⁸ <https://developer.amazon.com/en-US/docs/alexa/custom-skills/speech-synthesis-markup-language-ssml-reference.html>

To ensure that the data of different users is as comparable as possible, visual information is not displayed on Alexa devices that support this.

5.3. Study 3 (publication)

How to Design Audio-Gamification for Language Learning with Amazon Alexa?—A Long-Term Field Experiment

Paula Bräuer^a  and Athanasios Mazarakis^b 

^aComputer Science Department, Kiel University, Kiel, Germany; ^bZBW-Leibniz Information Centre for Economics, Kiel, Germany

ABSTRACT

Gamification can increase motivation in learning, and intelligent virtual assistants (IVAs) can support foreign language learning at home. However, there is a lack of design concepts to motivate learners to practice with their IVA. This study combines both concepts and analyzes if audio-gamification can increase engagement to address this research gap. To this end, a one-year long-term field experiment with 230 subjects using a German language learning skill for Amazon Alexa was conducted. A between-subjects design determined differences in learning behavior and learning outcomes between a control group and two gamified groups (achievements and leaderboard). The findings reveal a positive effect on the number of translated vocabulary and learning success. However, only in the group with a leaderboard was a statistically significant effect on the number of translated vocabulary found. These findings imply that audio-gamification can be a helpful tool for increasing motivation to use IVAs for foreign language learning.

1. Introduction

Individuals of many ages and backgrounds are interested in learning another language or may be obliged to do so (Young, 2014). However, learning a language takes time and effort, and language students often do not receive enough exposure or practice in the language they want to learn (Govender & Arnedo-Moreno, 2021). The development of new interaction methods with technological systems, such as intelligent virtual assistants (IVAs), can help learners with these challenges (Dizon, 2021; Istrate, 2019; Skidmore & Moore, 2019). In this respect, the term IVA is used interchangeably in the literature with terminologies, such as intelligent personal assistant (IPA), conversational agent (CA), conversational user interface (CUI), virtual personal assistant (VPA), or voice-enabled assistant (VA), to mention a few (Cowan et al., 2017). IVAs can be defined as “interfaces that enable users to interact with smart devices using spoken language in a natural way and provide a singular response to a query similar to speaking to a person” (McTear et al., 2016; Nobles et al., 2020).

IVAs like Amazon Alexa, Apple Siri, or Google Assistant are now available on various devices, such as smartphones and speakers. As a result, the use of speech to interact with automated systems has grown in popularity (Clark et al., 2019; Lopatovska et al., 2020). But despite their increasing prevalence, most individuals do not use IVAs regularly because they miss integration across devices and the ability to customize according to their regular usage habits (Cowan et al., 2017). However, regular practice and repetition are

essential to learn a new language. Therefore, for students to learn successfully with an IVA, student acceptance and satisfaction with the IVA are critical (Babic et al., 2018).

Moreover, IVAs are already used for learning new languages (Benner et al., 2022; Dizon, 2021; Istrate, 2019; Skidmore & Moore, 2019). By interacting with an IVA, for example, the listening comprehension, as well as the pronunciation, can be improved. Another advantage is that IVAs offer an easy way to train a new language interactively outside the usual classroom scenario (Istrate, 2019).

Even though the topic of IVA is increasing its relevance for research, there is currently a significant research gap in terms of design principles for IVAs (Clark et al., 2019). Qualitative studies, for example, show that many users rate IVA feedback often as inadequate and that feedback needs hints for improvement in a transparent way to achieve meaningful feedback (Lopatovska et al., 2020; Luger & Sellen, 2016). Thus, designing IVAs to encourage users to interact with them over time is challenging.

One approach that has been successfully applied in various contexts in Human–Computer Interaction to increase user motivation is gamification (Hamari et al., 2014; Koivisto & Hamari, 2019). According to Deterding et al. (2011), gamification is defined as “the use of game design elements in nongame contexts.” Besides the potential to increase user motivation, gamification has also been established as a helpful tool in the field of education (Bai et al., 2020; Dichev & Dicheva, 2017; Dicheva et al., 2015; Majuri et al., 2018). Aside from the most commonly observed

effects on motivation and performance, an increase in learning success has been repeatedly demonstrated in education with the help of gamification (Denny et al., 2018; Ortiz-Rojas et al., 2019).

Visual elements are common when designing and implementing a gamification approach; badges or leaderboards are typically used in a classroom setting (Christy & Fox, 2014; Sailer & Sailer, 2021; Tan & Hew, 2016). However, visual components should be avoided when integrating such game design elements into applications for an IVA since many end devices do not provide a screen, like Alexa Echo Dot or Google Nest Mini (Kinsella, 2019).

When designing dialogs, it is best to keep things as simple as possible in accordance with current design guidelines (Murad et al., 2018). However, integrating purely acoustically presented game design elements is a challenge because, due to fundamental differences in vision and hearing, elements from classic visual computer games cannot be transferred without considerable alterations to their properties (Friberg & Gärdenfors, 2004). Audio games, which, unlike traditional video games, do not have visual elements, are an example of how games can be created without a visual component (Garcia et al., 2013). Initial results have already shown that gamification can also be implemented purely acoustically without visual elements (Bräuer & Mazarakis, 2022), thus creating audio-gamification (Mazarakis, 2021). In a laboratory setting, Bräuer and Mazarakis found that the speed of processing household tasks was increased compared to a control group by audio-gamification.

As a research question, we aim to find support with this study that by combining the potentials of IVAs and audio-gamification in the field of language learning, positive effects on motivation and learning success can be achieved. In particular, unlike in previous studies (Bräuer & Mazarakis, 2022; Dizon, 2020; Moussalli & Cardoso, 2020; Sailer et al., 2017), the aim is not to work with students in a laboratory setting but to collect real-world data with a long-term field experiment that reflects real-world usage behavior. This article aims to contribute to this goal by investigating the integration of audio-implemented game design elements with IVAs. This approach is central because gamification studies usually rely on relatively short-term studies (Cermak-Sassenrath, 2019), and studies about learning with IVAs are typically conducted in lab settings (Dizon, 2020; Hsu et al., 2021; Pyae & Scifleet, 2018; Wu et al., 2020). By developing a dedicated application for Amazon Alexa, a so-called Alexa skill, it was possible to record usage behavior in detail in the various conditions. The evaluation of this quantitative data provides new insights that previous studies could not capture through surveys and self-reporting by users.

The rest of the article is organized as follows: The next chapter provides a brief theoretical background for the research. Then, the results of an empirical study are presented in the third chapter and discussed in the following. Next, limitations and future research approaches are referred. Finally, the results of the article are briefly summarized.

2. Background

IVAs in the context of language learning offer much potential for research (Dizon, 2021). At the same time, gamification in education is a popular topic in human-computer interaction (HCI) research (Dicheva et al., 2015; Seaborn & Fels, 2015). This section will review the related work in two aspects: language learning with virtual assistants and language learning in combination with gamification.

2.1. Learning foreign languages with intelligent virtual assistants

Many low-cost self-study tools exist to learn a foreign language, regardless of whether it is about reading or writing, but opportunities to practice speaking skills are far more limited (Ruan et al., 2021). IVAs can currently speak a variety of languages and can therefore be used as a language-learning tool. For example, Amazon Alexa is multilingual in nine languages (*Develop Skills in Multiple Languages | Alexa Skills Kit*, 2021), Google Assistant in 16 languages (*Change the language of Google Assistant—Android—Google Nest Help*, 2021), Siri is able to communicate in 21 different languages (*iOS and iPadOS-Feature Availability*, 2021), and the number of languages spoken by the assistants is constantly increasing. According to de Barcelos Silva et al.'s (2020) systematic literature review, education, in particular, could be a potentially relevant application area for IVAs, but it has not yet been sufficiently investigated.

Teaching with the help of IVAs poses many advantages. Speech recognition offers the possibility to train and specify the pronunciation of a newly learned language. Regular exchange in a foreign language with the IVA can also make learners aware of gaps in their linguistic knowledge (Dizon, 2020). In addition, meeting with a teacher to perform exercises is no longer necessary, but the student can train at home alone with the voice assistant (Istrate, 2019). So learning autonomy could be aided by IVAs, which provide learners with opportunities for language practice in a low-stress setting (Dizon, 2021). Also, most first-timers shy away from speaking in front of others when learning a new language (Namaghi et al., 2015). IVAs, on the other hand, could provide a friendly, non-intimidating environment for spoken language practice. In addition, the IVA's continuous availability also allows for responding to the learner's personal needs. The learner is not required to follow a predetermined lesson plan at a predetermined time but can practice with the IVA at any time of the day.

The scientific literature on IVAs in the context of language learning has focused on learners' attitudes and experiences with learning technologies. Several studies have reported difficulties for second language speakers (L2) related to being understood by an IVA (Chen et al., 2020; Dizon, 2017; Dizon et al., 2022; Pyae & Scifleet, 2019). Moussalli and Cardoso (2020) deal with this problem in more detail. Their results show that L2 learners have in contrast no problems understanding Alexa and that it adapts well to their accented speech. The results also show that L2 learners use a variety of strategies to address the

communication difficulties they experience with Alexa, e.g., repeating or rephrasing when problems in communication occur.

IVA interaction is often task-oriented and takes the form of questions and responses (Cowan et al., 2017). Vocabulary learning is an example of where this principle can be put to good use. Skidmore and Moore (2019) developed an Alexa skill called “Japanese Flashcards.” The authors consider areas of the Alexa development process that limit the facilitation of language learning, particularly the lack of multilingual speech recognition, and offer solutions to these limitations. The flashcard system applied in their article is commonly used for vocabulary learning. The concept is that the learner is presented with one side of the flashcard and tries to remember the content of the other side. Despite the fact that the Alexa skill has yet to be evaluated, the authors believe Alexa can contribute to the developing field of “voice-assisted language learning” by providing valuable learning paradigms, such as conversational role-play and pronunciation training (Skidmore & Moore, 2019).

In two similar studies, Ruan et al. (2019) compared a text-based chatbot app to a flashcard app. Even though both systems used the same algorithm to sequence the material, students recognized (and remembered) more correct answers when using the chatbot than when using the flashcard app. While using a chatbot was more time-consuming, in a second study, students spent 2.6 times more time on their own initiative using a chatbot rather than flashcards, indicating a strong preference for occasional learning. Also, the results of the second study revealed that chatbots outperformed flashcards in terms of learning gains in recall (Ruan et al., 2019).

Dizon (2020) investigated the effects of using Alexa on L2 English students in a quasi-experiment with an experimental group that received a 10-week treatment and a control group that did not receive such a treatment. The emphasis was on improving listening comprehension and speaking proficiency. According to the author, the experimental group was able to make more significant gains in L2 speaking proficiency. However, no significant difference was found when comparing improvements in L2 listening comprehension. The results of Dizon (2020) get support from Hsu et al. (2021), whose study also only showed an improvement in pronunciation but not in listening comprehension. Both studies used a setting where subjects performed various tasks with an IVA to practice a new language. For this purpose, the system language of the IVA was set to the foreign language to be learned. In addition, the subjects were given various instructions on how to interact with the device.

An important issue in working with IVAs for foreign language learning is the difference in communication behavior between native (L1) and L2 speakers. One finding by Wu et al. (2020) was that L2 users were seeking to lower their amounts of language production in conversation with the IVA. This was accompanied by frustration, having to reformulate queries from scratch. In contrast, L1 speakers tend

to pay more attention to the conciseness of their phrases to counteract the system limitations of the IVAs.

Pyae and Scifleet (2019) found similar results in their study, in which L2 speakers reported being able to easily learn to operate the IVA they were using but encountered considerable difficulty in obtaining the necessary words and formulating instructions in a way that the IVA understood. In addition, Wu et al. (2020) showed that L2 speakers favored smartphones because the visual feedback allowed for the diagnosis of communication problems while offering time to analyze the results. On the other hand, L1 speakers, who thought audio feedback was sufficient, chose smart speakers. Pyae and Scifleet (2018) showed in a study that the differences between L1 and L2 could also affect user experiences. In comparison, L1 speakers found IVAs easier and simpler to use and had more confidence in using them.

In a comprehensive literature review on IVAs for foreign language learning, Dizon (2021) identified opportunities and challenges in this area. A key finding of the author’s literature review was that many subjects in qualitative surveys complained that feedback in language learning through IVAs was very limited. Furthermore, there are instances where language is studied solely for testing purposes rather than for its communicative function or cultural richness. When students learn a language for its utility rather than its fun, they face obstacles, such as a lack of interest and motivation, which leads to a lack of practice (Govender & Arnedo-Moreno, 2021). By using various game design elements, gamification could assist in overcoming the problem of low motivation and generate new opportunities to provide users with relevant feedback (Mazarakis, 2015, 2021). To address this research and design gap, our study examines audio-gamification in the context of IVAs for foreign language learning.

2.2. Gamification in foreign language learning

Gamification has been defined as using game design elements in nongame contexts (Deterding et al., 2011). In learning, gamification is a design process in which game design elements are added to modify existing learning processes (Sailer & Homner, 2020). As a result, gamification became of significant interest to educators who have explored its potential to improve student learning (Bai et al., 2020; Dichev & Dicheva, 2017; Dicheva et al., 2015; Legaki et al., 2020; Majuri et al., 2018). In practice, there are also several examples of learning applications that use gamification. The best-known example is probably Duolingo (Adams, 2019), an application that can be used to learn a wide variety of foreign languages.

According to the theory of gamified learning, gamification influences learning outcomes by improving relevant attitudes and behaviors (Sailer & Homner, 2020). In addition, gamification is typically used to improve user motivation for a specific task or activity (Hamari et al., 2014; Sailer et al., 2017).

Motivational psychology theories and concepts have been particularly popular among HCI scientists for describing

and analyzing games and understanding what constitutes engaging player-computer interaction (Tyack & Mekler, 2020). In gamification research, Deci and Ryan's self-determination theory (Ryan & Deci, 2000) is commonly used to explain motivation (Seaborn & Fels, 2015). Motivation can be either extrinsic or intrinsic, according to this concept. Intrinsic motivation is characterized by doing something because it is inherently interesting or fun. On the other hand, extrinsic motivation refers to doing something because it leads to a definable result (Ryan & Deci, 2000, 2017). The theory of basic psychological needs is a derivative theory of self-determination theory (Deci & Ryan, 2000). According to this theory, there are three basic needs: experience of competence, a need for challenge, and feelings of effectance (Deci & Ryan, 2000); autonomy, a sense of volition or willingness when doing a task (Deci & Ryan, 2000); and social relatedness, which is experienced when a person feels connected with others (La Guardia et al., 2000). The satisfaction of these three needs might lead to intrinsic motivation. The concept argues that an individual's environment has an impact on their satisfaction with basic needs. This influence can be both positive and negative, contributing to or limiting the satisfaction of needs (Vansteenkiste & Ryan, 2013). However, it can be assumed that a significant part of the motivation in daily tasks is extrinsic and that gamification is externally endorsed (Mazarakis & Bräuer, 2022; Mekler et al., 2017). Nevertheless, it is not impossible that in some cases, an extrinsic motivation could be converted into an intrinsic motivator.

A literature review by Majuri et al. (2018) shows that gamification in the educational context achieved similar results as in other contexts. Most of the studies examined psychological aspects of the effect of gamification and mainly reported positive results. In addition, most studies found a positive effect on measurable educational outcomes. There are a variety of metrics to measure such an effect (Looyestyn et al., 2017), including, but not limited to course grades (Denny et al., 2018; Dichev & Dicheva, 2017), time spent (Landers & Landers, 2014; Monterrat et al., 2015), or volume of contributions (Bouchrika et al., 2021; Tejedor-Garcia et al., 2020). Since successful gamification involves repetition of the desired results (Robson et al., 2015), it makes perfect sense to apply it to vocabulary learning, where repetition is also essential to improve recall.

Dehghanzadeh et al. (2021) conducted a literature review on gamification in the context of learning English as a foreign language, which included 22 publications. Besides grammar or pronunciation training, vocabulary learning was by far the most common application in which gamification was used in the studies considered. None of the publications considered there could show negative results from the integration of gamification. Slightly more than half of the studies stated that the learners' experience was positively influenced by the integration of the game design elements. In terms of the literature reviewed, the authors criticize, among other things, the fact that many of the studies were conducted without a proper control group and that the data

was mostly collected in a very short period of time, which can be considered as important research gaps.

How gamification can be used to enhance IVA-supported learning of English as a foreign language was investigated in a study by Tejedor-Garcia et al. (2020). The authors report the development and evaluation of a gamified application for practicing English vocabulary pronunciation. For voice recognition, an app was created for smartphones that interact with Google Assistant. To encourage users to stay motivated, challenges were developed in which they could compete. A comparison was made with data from an earlier study (Tejedor-Garcia et al., 2016) to see how the game design elements affected motivation and performance in pronunciation practice. Learning success and motivation exhibited positive trends in the comparison. Visual representation was employed to implement the game design elements (challenge, leaderboard, points, avatar, and badges). The results suggest that gamification is helpful in this context. However, it remains an open question whether gamification would have a positive effect without visual components.

So far, it is unclear which game design elements should be used, both for IVAs (Bräuer & Mazarakis, 2022) and in general (Mazarakis & Bräuer, 2022; Mekler et al., 2017). A general analysis of game design elements used in the context of digital game-based language learning programs was conducted by Govender and Arnedo-Moreno (2021). The most frequently analyzed programs were those for vocabulary learning, considering different age groups and target languages. According to the analysis, the most frequently used game design elements were feedback, theme, points, narrative, and level. The authors note, however, that generalizing results is usually difficult because no attention is paid to specific game design elements or their combinations. This problem is already known and researchers are looking for a systematic examination of individual game design elements to draw more precise conclusions about their effects (Mazarakis & Bräuer, 2018, 2022; Mekler et al., 2017).

There are numerous game design elements used in gamification research and practice, but the majority of gamification approaches are developed on a foundation of three game design elements: points, badges, and leaderboards (Huang et al., 2020; Seaborn & Fels, 2015; Werbach & Hunter, 2012). Based on the predominantly positive results for these three game design elements in many other studies, it was decided to investigate the effect of these elements in our study as well.

Points are most often the basis for other game design elements, as they can be used to capture and store various game metrics (Werbach & Hunter, 2012). They serve as a numerical unit to indicate progress (Seaborn & Fels, 2015) or as a numerical representation of rewards for performing certain activities (Vitkauskaitė & Gatautis, 2018). Unlike binary feedback (right or wrong), points can be used to indicate how close an answer is to the optimal solution and are therefore often used in learning systems (Kapp, 2012).

Badges and achievements are a form of feedback awarded to the user for completing certain tasks (Marczewski, 2018).

Badges are typically viewed as a visual representation of achievements (Antin & Churchill, 2011; Groening & Binnewies, 2019; Werbach & Hunter, 2012). Because badges, as previously noted, are among the most studied game design elements (Huang et al., 2020; Seaborn & Fels, 2015), they are considered in this study. At the same time, however, visual components should be avoided in the context of IVAs. Since the two terms are often used synonymously in the literature (Werbach & Hunter, 2012) and the effect of badges and achievements are comparable, the decision was made to use achievements in the study. According to Antin and Churchill (2011), achievements serve five functions: setting goals, giving instructions, allowing reputation to be valued, creating a group feeling, or serving as a status symbol. Groening and Binnewies (2019) examine the effect of achievements in detail. Their results show that achievements can increase performance and motivation. However, as with many other game design elements, the effect depends strongly on the design. Based on the results of their study, the authors recommend using achievements with a high degree of difficulty and in small numbers.

A leaderboard is another game design element and consists of an ordered display of scores against a specific success criterion, along with the names of users (Groening & Binnewies, 2019). Its purpose is to make simple comparisons (Zichermann & Cunningham, 2011), especially addressing the basic need for social relatedness (Bräuer & Mazarakis, 2019; Sailer et al., 2017).

Ortiz-Rojas et al. (2019) integrated a leaderboard into an engineering course and compared learning performance, intrinsic motivation, self-efficacy, and students' engagement with a control group that did not have a leaderboard over a four-week period. To assess learning performance, the results of a knowledge test were used. The other three variables were assessed using a questionnaire, which was completed at the start and end of the study. The integration of the leaderboard had a positive effect on learning performance, according to the findings. On the other hand, the other three factors did not become statistically significant. Finally, Landers and Landers (2014) investigated the impact of a leaderboard in the context of a course project. The authors assigned students to a leaderboard randomly and discovered that the leaderboard was associated with more time spent working on the project and, as a result, better project performance.

3. Method

After presenting the related literature, we now detail the methods in this chapter. The study examines the impact of audio-gamification on motivation and learning performance and thus investigates the integration of audio-implemented game design elements with IVAs. The hypotheses and experimental design are described in the following section, including a detailed description of the data cleaning for the statistical analysis.

3.1. Hypotheses

Based on previous findings in the scientific literature, we developed two hypotheses for the experiment, each split into comparisons between the control group and an experimental group. These are measured as the number of vocabulary processed, and learning success is measured as the number of words correctly reproduced.

The duration over which the subjects interacted with the experiment, more precisely with the Alexa skill, was analyzed to investigate the effect of the game design elements on user motivation. Because previous research has shown that gamification can increase general engagement (Landers & Landers, 2014; Majuri et al., 2018; Monterrat et al., 2015), and gamification encourages learners to engage more deeply with their learning material (Bouchrika et al., 2021; Dichev & Dicheva, 2017), the number of vocabulary processed was analyzed to determine whether the incorporation of game design elements influences learning motivation. As a result, we research the following two hypotheses:

H1a: The group with the leaderboard condition processed more vocabulary than the control group, with no game design elements.

H1b: The group with achievements condition processed more vocabulary than the control group, with no game design elements.

Second, the impact of the game design elements on subjects' learning performance will be investigated. Again, we refer to the literature results presented earlier (Denny et al., 2018). The ratio of vocabulary correctly processed is analyzed to measure learning success. Two additional hypotheses are examined:

H2a: The group with the leaderboard condition has a higher ratio of correct vocabulary than the control group, with no game design elements.

H2b: The group with the achievements condition has a higher ratio of correct vocabulary than the control group, with no game design elements.

3.2. Experimental design

A skill was developed for the Amazon Alexa platform that allows learning German as a foreign language to collect the data for this study. Because the speech input is determined by Alexa's default system language, this means that an IVA with an English setting as a system language cannot process German-language input. However, with the help of Amazon Alexa's Speech Synthesis Markup Language, it is also possible to reproduce texts in other languages *via* text-to-speech (*Speech Synthesis Markup Language (SSML) Reference | Alexa Skills Kit, n.d.*). Therefore, the Alexa skill is designed to train listening comprehension, according to the flashcard concept that Skidmore and Moore (2019) describe in their article. In addition to the advantage of not requiring users to make any additional system settings, this approach

provides the benefit of assuming that native speakers can easily interact with the IVA (Wu et al., 2020).

To ensure that a long-term effect can be demonstrated, more than 500 vocabularies were selected and integrated into the Alexa skill so that the users are always provided with new vocabularies, even when using the Alexa skill intensively. Of the words, 70% were nouns, 20% were verbs, and 10% were adjectives. In addition, many words regularly used in everyday language were chosen to make the Alexa skill enjoyable to beginners. Looking at the corresponding word frequency level, 38% of the words in the Alexa skill correspond to level 1, 25% to level 2, and only 4% to level 3. The remaining 33% of words could not be classified *via* the “Frequency Level Checker” tool to determine the word frequency level (Maeda, n.d.).

Some terms were frequently misunderstood when conducting pre-tests and had to be excluded. For example, “to buy” was misinterpreted as “goodbye,” resulting in the Alexa skill being mistakenly ended in numerous incidents. However, the issue is widely recognized (Wu et al., 2020), and developers should always guarantee that every expected user input is unique to reduce user annoyance. This challenge is amplified when using Alexa skills to learn foreign languages because of the imprecise pronunciation of L2 learning (Wu et al., 2020). Nevertheless, the difficulty was mitigated in our study since the Alexa skill trains listening comprehension rather than pronunciation. Furthermore, the speech input was given in the trainees’ native language, which in our instance was English. Still, Alexa may misinterpret even L1 speakers. Therefore, we counted how often each word was classified as correct or incorrect in the pre-tests. As a result, suspicious words were studied in-depth and replaced with less error-prone words.

Subsequently, to check the error susceptibility of the Alexa skill, an evaluation of the Automatic Speech Recognition (ASR) was performed using the test tool provided by Amazon. Using a test set created *via* the Alexa Developer Console, consisting of 762 test cases, a value of 75% was achieved. The ASR evaluation tool can be used to test audio samples with ASR models and compare the expected transcriptions with the actual transcriptions. The ASR analysis thus shows, for example, when “hettie” is understood instead of “headache.” In addition, an intent confidence by utterance of 100% can be reported for the study period. No case was recorded with a low confidence, so according to the analysis, a proper answer was always found for the user’s input.

To be able to investigate the influence of audio game design elements in a controlled setting, a 3×2 between-subjects design was implemented in which the control group was compared with each experimental group. Thus, the experimental setting includes the following groups:

- The control group without any game design element
- An experimental gamified group with a leaderboard
- An experimental gamified group with achievements

To acquire as many subjects as possible for the study, the Alexa skill was made available in all English language

variants in the Amazon Store (US/CA/AU/IN/UK). Log files were used to record how often individual subjects used the Alexa skill, how long they interacted with it, how many vocabulary words were learned per session, and the success rate.

The experimental design of the Alexa skill differs among the groups. The two experimental conditions were each enhanced by the group-specific game design elements and functionalities to support gamification. The control group did not have any of these functionalities.

When subjects start the experiment for the first time, they are randomly assigned to one of the three conditions and cannot switch between the groups. They are informed that the Alexa skill is being used in a research project and that data will be stored for this purpose. The subject was also informed where to obtain more information about the Alexa skill and the research project. In addition, it was highlighted that the participation was entirely voluntary and that the deletion of the account and the associated data is possible at any time. The subject’s anonymity was guaranteed by not asking for any additional information. This procedure is identical in all three groups. Any variations are described accordingly. In addition, section 3.3 provides a more detailed description of how and when the data were collected. After the introduction, vocabulary training begins. Figure 1 illustrates this procedure.

In the following, we first describe how the control group was designed and then go into more detail about the design of the individual game design elements for the two experimental conditions.

3.2.1. Design of the control group

In the control group, the vocabulary was asked on each trial until the subject answered all vocabulary or quit the Alexa skill whenever they wanted. The order in which the vocabulary is asked corresponds to the levels used in both other groups, which are explained in more detail in the next section. After each subject’s response, either the correctness of the given answer is validated with positive oral feedback (e.g., “this is correct”) or negative oral feedback, and the proper translation is provided. If a user has processed the entire vocabulary, learning will start from the beginning. The process is illustrated with a simplified flow chart in Figure 2. Unless the Alexa skill does not understand the user because (a) no input is provided or (b) an answer is given that the Alexa skill does not understand, Alexa apologizes and requests the answer to be repeated. This is repeated twice. If the Alexa skill does not receive an answer after the second request, Alexa switches off.

3.2.2. Design of the gamified groups

Both gamified groups (achievements and leaderboard) differ from the control group in some aspects. These are described in this section and additionally illustrated at the end of this section in Figure 3. In contrast to the control group, after informing the user that the data is being collected for research purposes, the subjects are asked for their nickname

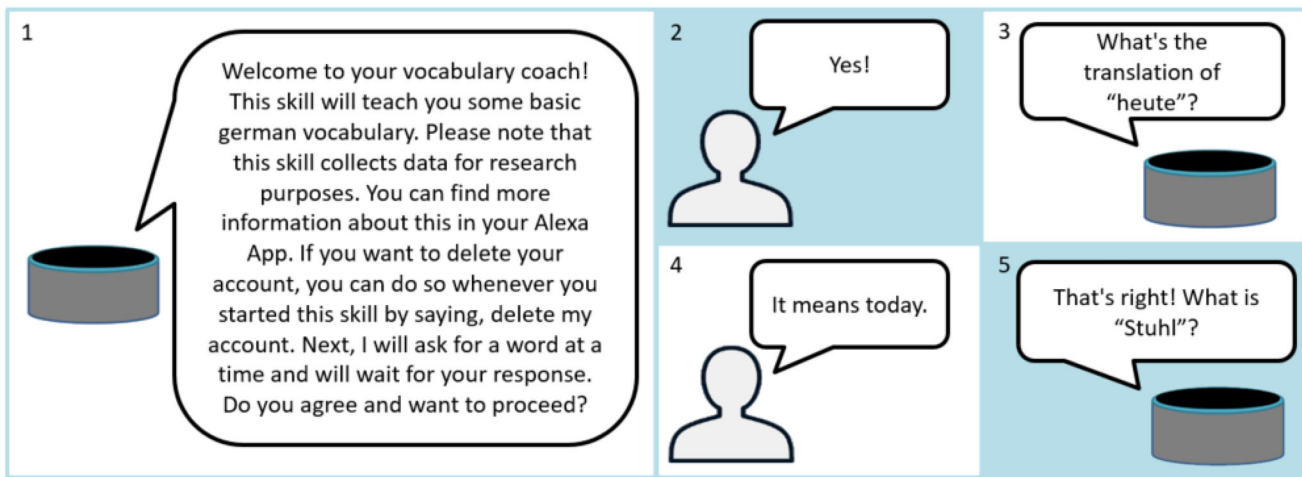


Figure 1. Example dialog in the control group when starting the Alexa skill for the first time.

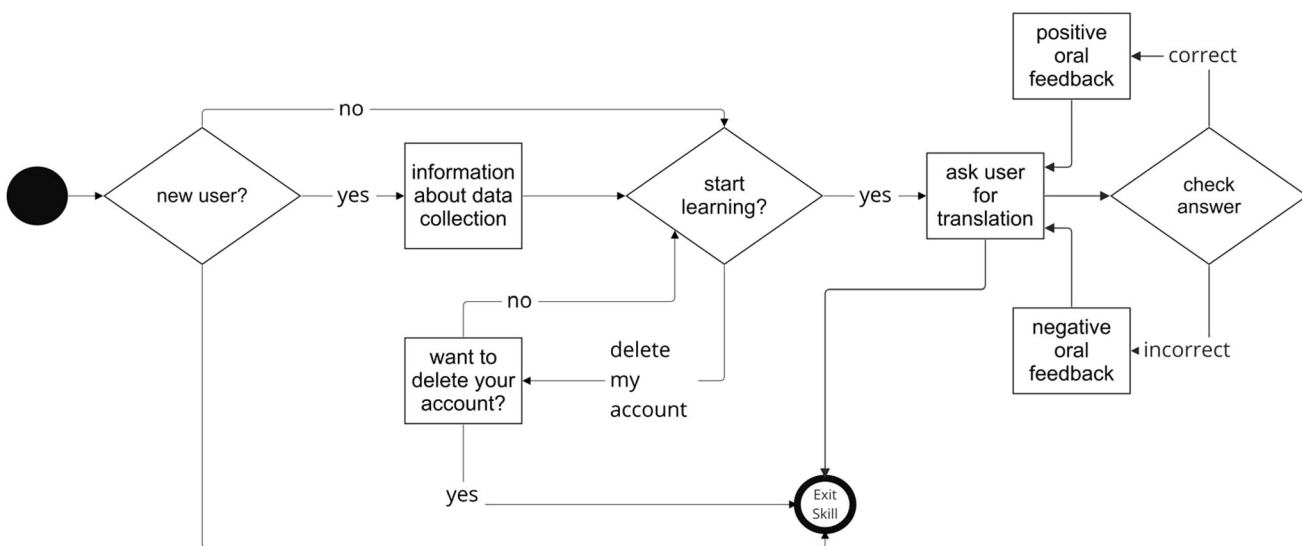


Figure 2. The flowchart illustrates the interaction with the Alexa skill in the control group.

at the first start. The nickname is used for a personal greeting when the experiment is accessed again. It is not necessary for the subject to state their real name; any nickname is acceptable. Now the user can choose between five topics: everyday life, health, leisure, nature, and education. The decision leads to the presentation of vocabulary from the chosen topic. In addition, the subjects are informed of their progress in percent. To avoid overwhelming the subjects with too much information and based on pre-test findings, the information about the progress only is given with a probability of 20%, so generally speaking, on average, every fifth time. Depending on the group, with leaderboard or achievements, the subjects also get information about the features of the respective game design element (see Figures 4 and 5).

The five topics are divided into 13 different levels, each topic consisting of 100 vocabulary words. The first six levels contain five different vocabulary words, and from the seventh level onwards, ten different vocabulary words are requested.

The subjects begin with the training session identical to the control group's vocabulary retrieval procedure, i.e., vocabulary is learned. After all vocabulary from a level has been asked and correctly answered at least once, the subject can practice the level again or take a challenge. During a challenge, all vocabulary from the corresponding level are randomly asked again.

When starting a challenge, the user receives life points. Each incorrect answer costs the subject one of their life points. If they run out of life points and the challenge has not been successfully completed yet, they lose and can restart the challenge from the beginning. If the subject chooses not to complete the challenge or fail it, there is the option to return to the main menu. Lower levels (1–6) begin with three life points, while higher levels (7–13) begin with four. After a challenge has been passed or failed, a corresponding audio feedback is heard, which is intended to emphasize the result of the challenge. An example of the dialog sequence in which a challenge is started is shown later in Figure 5.

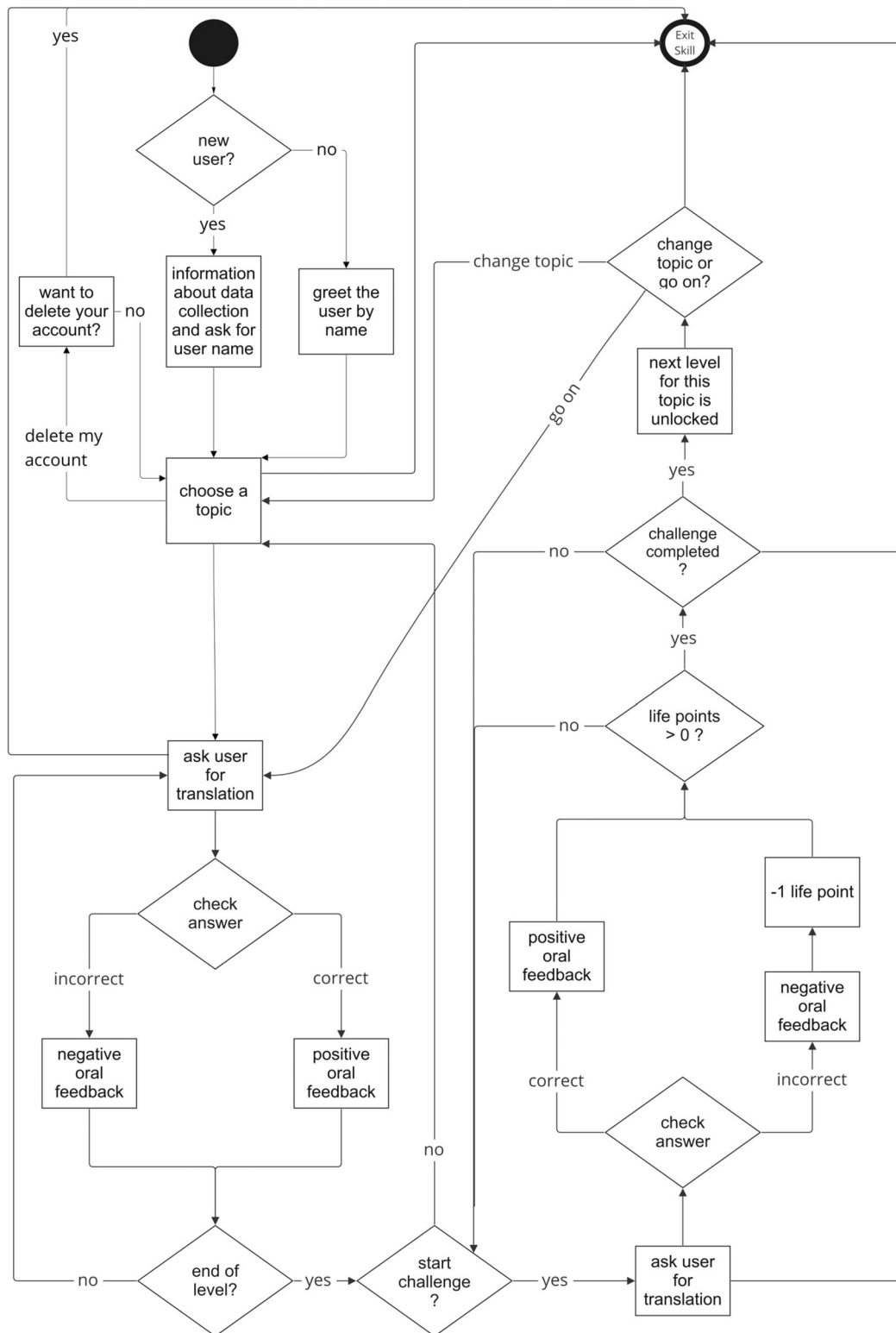


Figure 3. General flowchart illustrates the interaction with the Alexa skill in the experimental groups (achievements and leaderboard).

Once a challenge has been successfully completed, the next level for the corresponding topic is unlocked, and positive audio feedback is played. The subject may proceed directly to the next level's training unit or return to the main menu to change the topic. After completing all levels of a topic, the user gains access to a final challenge in which all vocabulary from all levels of the area is randomly asked.

3.2.3. Design of the leaderboard

The design for the first experimental group, the leaderboard condition, differs from the other groups. When the Alexa skill is activated, the subjects are told their position on the leaderboard with a 20% likelihood, according to the pre-test findings, to avoid overwhelming the subjects with too much information.

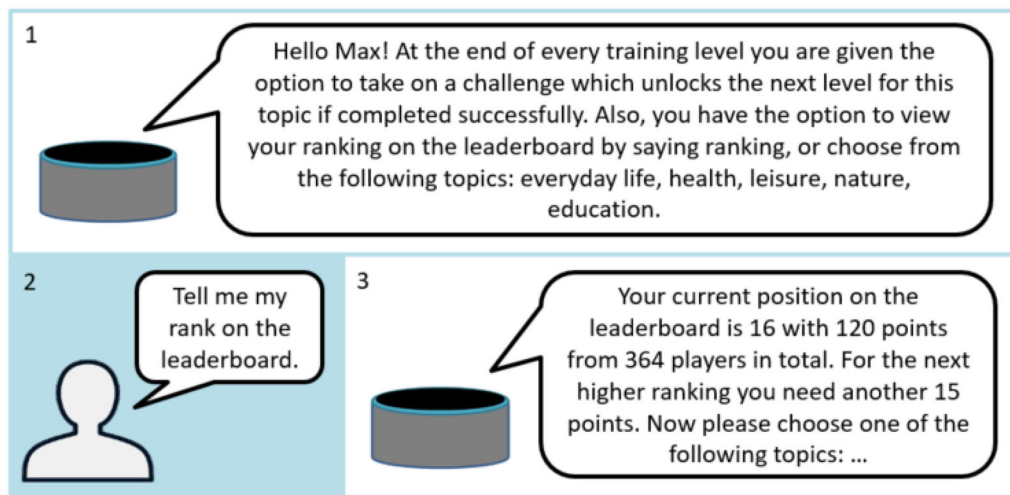


Figure 4. Example of a dialog in which the user calls up the leaderboard.

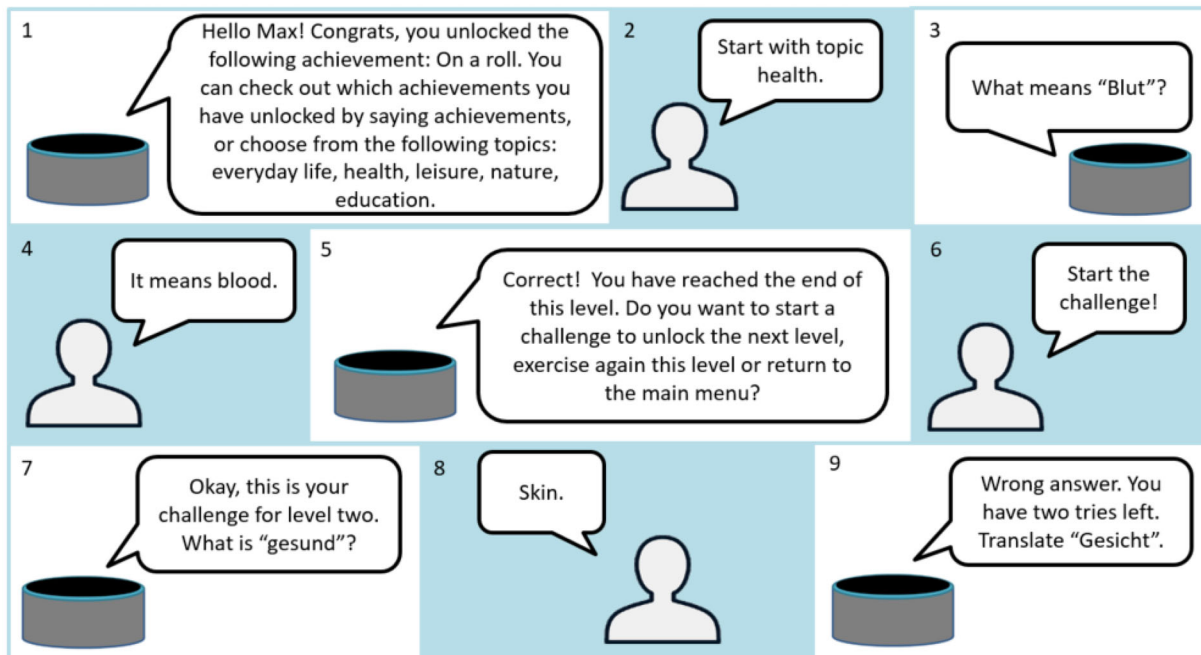


Figure 5. Example dialog for a user who receives an achievement and starts a challenge.

The subjects in this condition collect points during the challenges to move up in rank. The number of points awarded for a completed challenge depends on three factors: the level, whether a question was answered correctly, and how many times the challenge was started. For example, for levels 1–6, 125 points for each correct answer were awarded for the first attempt, 25 for the second, and five for the third and further attempts. The corresponding points for levels 7–13 were 250, 50, and 10. The staggered scoring is intended to encourage the user to complete a challenge in as few attempts and as good as possible, thus expediting their progress. Additionally, this method of awarding points is intended to prevent users from cheating. For example, a user earns no points by intentionally aborting and restarting the same challenge. An example of how the leaderboard is communicated to the user is shown in Figure 4.

3.2.4. Design of the achievements

Users in the second experimental group, the achievement condition, can unlock eleven achievements under certain conditions, as listed in Table 1. According to Werbach and Hunter (2012), badges or achievements can be used to assist in onboarding a new system. Badges should therefore be easy to reach right from the start so that users are encouraged to continue interacting with the application (Kapp, 2012). Therefore, the design of the achievements tries to give the user the possibility to unlock them already at the beginning, e.g., for completing a challenge for the first time or even giving a wrong answer. Another criterion for design is based on the results of Groening and Binnewies (2019). Therefore, the number of achievements was limited and was designed to be more challenging, such as working through a whole topic or reaching a final challenge.

Table 1. List of all achievements with their title and description.

| Achievement | Description |
|------------------------------------|--|
| Getting started | The user has completed the first challenge. |
| Every beginning is difficult | The user has failed a challenge. |
| Perfect one | The challenge of a ten-word level was completed without a mistake. |
| Halfway there | The user has completed fifty percent of all levels. |
| Topic master | The user has completed all the final challenges. |
| On a roll | Unlocked when the user has used the Alexa skill for three consecutive days. |
| Ambitious | The user has started the Alexa skill 20 times. |
| Practice makes perfect | Before starting a challenge, the user took the chance to practice the level again. |
| Time for the real challenge | The user has unlocked the first final challenge. |
| That is not what I meant, but okay | The user has used a registered synonym as an answer during training or during a challenge. |
| You are the vocabulary coach | The user has unlocked all the previous achievements. |

An announcement informs subjects when an achievement is unlocked. These occur based on the state of the achievement at the end of a challenge, after answering a question, or when starting the Alexa skill. In addition, when an achievement is unlocked, audio feedback is played. The subjects can have their unlocked achievements read aloud to them from the main menu. The title of the achievement and its meaning are briefly explained. The user is then returned to the topic selection. An example of a dialog in which the user receives an achievement when starting the Alexa skill is shown in Figure 5.

3.3. Data collection

The underlying data for the results were collected in a long-term field experiment for one year, more precisely, between August 2020 and August 2021. For this purpose, the necessary Alexa skill was made freely available for download in the Amazon Store. The Alexa skill and the experiment were not advertised. Any participation was based on the random download of the Alexa skill from the store. During the experiment period, the Alexa skill was used by 488 unique users. No payment was offered or paid to the subjects.

Users were informed that their data was being gathered for research reasons on the Alexa skill's webpage in the Amazon Store and the Alexa app, as well as when they first started the Alexa skill. Each skill uploaded to the Amazon store automatically gets a webpage where users can see more information about that skill. This page can be found on the Amazon market, similar to pages for products sold on the platform. Our Alexa skill page explains what information is stored, such as the nickname and interaction logs. In addition to the website, each skill receives an entry in the Alexa app. The app is used to set up and operate Alexa devices. When a user activates a new Alexa skill, it appears in the app. Here, too, information about the research project and instructions on how to delete the collected data were provided for our Alexa skill. In addition, when starting the Alexa skill, subjects were told that data is stored for research purposes and that more detailed information is available in the Alexa app.

Before the subjects could continue interacting with the Alexa skill, they had to give consent by saying "yes" to the fact that they understood this information and wanted to

continue using it. If a subject does not agree with this requirement at any time during the experiment, it is possible for them to delete the collected data *via* voice command. Because this step cannot be reversed, the user is asked to confirm this action twice before proceeding. It is possible to restart the experiment after deleting the data. However, the subject is treated as a new user and may be assigned to a different group due to the randomized assignment. The delete function was used four times throughout the experiment.

For the experiment, log files were collected that documented the usage behavior of the Alexa skill. For each user, an entry was created in a database when the Alexa skill was started for the first time. Interactions with the Alexa skill were then recorded, and the corresponding information, such as the correct or incorrect answer to a vocabulary word, was stored in the database. All data collected is anonymized, and user re-identification is not possible. To evaluate the hypotheses, the following general data were collected: ID of the user, group assignment, start and end time of a session, and frequency of use of the Alexa skill. The frequency of use increases with each access to the Alexa skill. The usage time is calculated from the start and end time entries in the database. In addition, the total number of vocabulary processed and the proportion of correctly and incorrectly answered vocabulary are stored.

In both gamified groups, additional information was stored to implement the game design elements. These include the nickname of the user, which level in which topic the user has reached, the total percentage of progress, the current score, the number of times the leaderboard has been accessed, and the current number of times a challenge of a specific level has been taken. In addition, for the group with achievements, the unlocked achievements are saved with the timestamp when they got unlocked and the number of times the achievement overview has been accessed.

3.4. Participants

Not all data collected during the study period can be used for the subsequent analysis of the hypotheses. This chapter, therefore, describes the procedure used to exclude data sets that were not usable for the analysis. Subsequently, a

Table 2. Mean, standard derivation, and median of the sum of processed vocabulary, correct answers, and time in seconds per user.

| | Sum vocabulary | | | Correct vocabulary | | | Time in seconds | | |
|--------------|----------------|-----------|------------|--------------------|-----------|------------|-----------------|-----------|------------|
| | <i>M</i> | <i>SD</i> | <i>Mdn</i> | <i>M</i> | <i>SD</i> | <i>Mdn</i> | <i>M</i> | <i>SD</i> | <i>Mdn</i> |
| Control | 24.68 | 36.85 | 9.50 | 11.30 | 20.82 | 2.00 | 287.83 | 367.97 | 144.50 |
| Achievements | 25.79 | 30.49 | 14.50 | 13.23 | 17.45 | 7.00 | 418.58 | 393.96 | 275.00 |
| Leaderboard | 26.87 | 34.44 | 16.50 | 16.12 | 25.15 | 9.00 | 392.14 | 379.29 | 267.50 |
| Total | 25.86 | 34.13 | 13.00 | 13.75 | 21.88 | 5.00 | 364.80 | 381.98 | 223.50 |

description of the subjects is summarized based on the cleaned data.

3.4.1. Data cleaning

Data cleaning is performed based on two data sets: The sessions conducted and the subjects. A session has a starting point when individuals start the Alexa Skill and an endpoint when they stop using the Alexa skill, intentionally or unintentionally. A new session is created when a user invokes the Alexa skill a second time. To merge the data of the individual sessions, a unique user ID generated by Amazon is used. This unique key remains the same even if the user calls up the Alexa skill again so that the assignment to the conditions remains the same and the progress in learning the vocabulary can be documented. Over the entire period of the study, data was collected on 488 subjects (156 in the control group, 152 in the achievements group, and 180 in the leaderboard condition) with associated data on 941 sessions (individual access to the Alexa skill).

First, from 488 unique user IDs in the dataset, 229 entries were deleted because not a single vocabulary was processed over all sessions. In addition, another 17 subjects who had listened to only one vocabulary word but answered it incorrectly were removed. In the next step, one subject was removed from the analysis because the behavior was characteristic of a bot and not a human (access always took place at the same time, and none of the asked vocabulary words were answered correctly over 84 sessions). Finally, another subject was excluded from the dataset because out of 83 vocabulary words, zero were answered correctly. After these steps, entries from 240 subjects who interacted with the Alexa skill in 375 sessions remain.

Finally, outliers were excluded from the dataset because they studied a disproportionately large number of vocabulary items and appeared to be highly intrinsically motivated. The mean value of the processed vocabulary for all users (40.70) was calculated for this purpose, and the double standard deviation (2×85.40) was added. This led to removing ten more users from the dataset who edited more than 211.50 vocabulary words. The final sample for analysis consists of 230 subjects and 289 combined sessions.

3.4.2. Description of the final sample

The 230 subjects are distributed among the three experimental conditions, with 76 in the control group, 62 in the achievements group, and 92 in the leaderboard group. Besides, the Alexa Developer Console revealed which language settings were used on the devices by users: 24% from the United Kingdom, 5% from Canada, 46% from the

United States, 2% from Australia, and 23% from India. While this enables conclusions about users' locations to be drawn, it is not the same as geolocations. For example, while a user in Canada is more likely to set the language to Canadian English, they could just as easily set it to British English or even German for their end device.

4. Results

This section provides the descriptive and inferential statistical results. This includes the effect of gamification on usage behavior and on learning success. Finally, further statistical analysis of whether the assumptions regarding the time duration are valid is conducted.

4.1. Descriptive results

In the control group, the experiment was started on average 1.29 times per user (*SD* 0.78), in the group with achievements 1.35 times per user (*SD* 0.89), and in the group with leaderboards 1.16 times per user (*SD* 0.50). So for repeated Alexa skill usage, no noticeable statistical differences or biases exist between the groups.

To answer the hypotheses, the means, standard deviations, and medians for the total vocabulary processed, the percentage of vocabulary words answered correctly, and the time in seconds that users interacted with the Alexa skill are summarized in Table 2.

4.2. Effect of gamification on usage behavior

To evaluate the hypotheses, we first examine how the game design elements affect the number of vocabulary processed. The data are first tested for normal distribution before comparing the mean values between the control and experimental groups. The Shapiro-Wilk test produces a statistically significant result, $p < .001$. As a result, the Mann-Whitney *U* test is used as a non-parametric statistical method for comparing mean values between groups (Field, 2017).

The comparison between the control group and the experimental group with achievements is not statistically significant: $U = 2002$, $z = -1.52$, $p = .065$ (one-tailed), $r = .13$. As a result, hypothesis H1a cannot be supported, and it cannot be assumed that achievements have a positive impact on learners' motivation to engage more deeply with their learning material.

The comparison between the control group and the group with the leaderboard yields a statistically significant result: $U = 2873$, $z = -1.99$, $p = .024$ (one-tailed), with a small effect size $r = .15$. The effect size for the hypothesis

Table 3. Mean, standard derivation, and median ratio of correct vocabulary to the sum of answered vocabulary.

| | <i>M</i> | <i>SD</i> | <i>Mdn</i> |
|--------------|----------|-----------|------------|
| Control | 0.300 | 0.229 | 0.333 |
| Achievements | 0.398 | 0.265 | 0.417 |
| Leaderboard | 0.457 | 0.253 | 0.500 |
| Total | 0.390 | 0.257 | 0.417 |

H1b is interpreted according to Field (2017). Thus, hypothesis H1b can be supported, and it can be assumed that adding a leaderboard increases the motivation to practice more vocabulary with the Alexa skill.

4.3. Effect of gamification on learning success

The ratio of correct vocabulary to the sum of answered vocabulary was first calculated to investigate the impact of game design elements on learning success. This is necessary to answer the hypotheses H2a and H2b. Table 3 summarizes the findings. For hypotheses H2a and H2b, the interpretation of the effect size is made according to the benchmarks provided by Plonsky and Oswald (2014) for L2 research.

The sample was tested once more for normal distribution for further analysis. The Shapiro-Wilk test also becomes statistically significant here, with $p < .001$. The comparison between the control group and the experimental group with achievements becomes statistically significant, $U = 1829.5$, $z = -2.27$, $p = .012$ (one-tailed), with a small effect size of $r = .19$. In L2 research, a correlation coefficient interpreted as effect size has a small effect size around 0.25 and a medium effect size around 0.40, leaving a large effect size around 0.60 (Plonsky & Oswald, 2014). As a result, Hypothesis H2a can be supported, and it can be assumed that achievements have a positive influence on learning success.

The comparison between the control group and the experimental group with a leaderboard gives a statistically significant result as well, $U = 2223.5$, $z = -4.07$, $p = .000$ (one-tailed), with a small effect size of $r = .31$. Thus, the hypothesis H2b can also be confirmed, and it is assumed that leaderboards also positively influence learning success.

4.4. Further analyses

The time users spend with the Alexa skill per session is another aspect that could yield further interesting insights about the influence of audio-gamification. Since the dialogues in the gamified groups are much longer than in the control group (as can be seen by comparing Figures 2 and 3), no firm conclusions about learning behavior can be established. However, time per session may still provide insights into user behavior. In line with the previous hypotheses, it is assumed that in the gamified conditions, there are increases compared to the control group. Following that, we examine how the game design elements affected the interaction time of use in seconds. Again, we first check the normal distribution of the data and obtain a statistically significant result using the Shapiro-Wilk test, $p < .001$. As before, we need to use the Mann-Whitney U

test, a non-parametric statistical method for comparing mean values between groups (Field, 2017).

The comparison between the control group and the experimental group with achievements becomes statistically significant, $U = 1495$, $z = -3.68$, $p = .000$ (one-tailed), with a medium effect size $r = .31$. Furthermore, the comparison between the control group and the experimental group with the leaderboard condition provides a statistically significant result, $U = 2257$, $z = -3.95$, $p = .000$ (one-tailed), also with a medium effect size $r = .30$.

Prior knowledge of the subjects could bring a bias into the data in such a field experiment, which could significantly complicate the interpretation of the results. For reasons of data economy and following the privacy by design principle, we did not ask for this information in our study. However, the analysis of the existing data is sufficient to exclude such a bias. The main idea for identifying such a bias lies in the skewness and distribution of data. Despite the premise that, in general, a sample size of 30 and more subjects, especially if it is the case in each group, is widely accepted as large enough (Field, 2017), we can still analyze the data further to find irregularities about learning behavior. For this reason, we further analyzed the ratio of correct vocabulary. First, all students with zero correct answers were removed, as prior knowledge does not exist for them and cannot lead to a bias. The number of subjects removed is almost equally distributed across all three groups (20 for the control group and 12 each for the achievements and leaderboard groups). Next, it was analyzed if the data distribution and skewness of the data for the remaining subjects in each group were normal. This is indeed the case: The skewness for the control, achievements, and leaderboard groups are 0.024 (SD 0.319), -0.102 (SD 0.337), and -0.277 (SD 0.269), all far away from a critical value of 1 or -1 (Field, 2017). In addition, the Shapiro-Wilk test is not statistically significant for any group, with $p = .131$ for the control group, $p = .424$ for the achievements group, and $p = .235$ for the leaderboard group. These results indicate that there is no apparent bias in the groups and that the participants have no obvious bias in their learning behavior. Of course, this is not strong statistical evidence that the participants are similar in terms of learning attitudes. Nevertheless, it is still a powerful indicator that the subjects in each group do not differ in learning behavior, which could be the case if participants in one group would benefit from prior knowledge.

5. Discussion, limitations, and future research

The motivation for this long-term field experiment was to advance the field of conversational user interfaces with the combination of incentives to use and so to enhance language learning with an IVA. Therefore, we systematically defined, analyzed, and compared the use of audio-gamification for one year with a field experiment on practicing listening comprehension of German vocabulary with an Amazon Alexa skill.

What distinguishes this study from previous research in both (a) IVAs for language learning (Dizon, 2021) and (b)

gamification for language learning (Dehghanzadeh et al., 2021; Huang et al., 2020; Seaborn & Fels, 2015) is the combination of IVAs and gamification as well as the collection of data *via* a long-term field experiment. By conducting the study in a natural non-school setting, i.e., informal and non-guided, realistic findings on the use of IVAs as a tool for L2 learning could be collected (Dizon, 2021) and thus achieve a high external validity. At the same time, in contrast to many other studies (Dehghanzadeh et al., 2021), this article also reports the results of a control group, which distinguishes it from many other studies. In line with previous literature on gamification, we find positive effects but also mixed results.

Regarding the first hypothesis, a statistically significant influence on the number of processed vocabulary words was found using the leaderboard but not for the achievements condition. As already briefly described, in contrast to the visual representation of gamification, every game design element in audio-gamification must be represented by sound or speech. Consequently, information about unlocking achievements cannot be displayed as, e.g., a small element in the upper right corner of a screen. Instead, the actual activity must be interrupted for the game design element to be expressed. As a result, the game design elements are at odds with increasing motivation and distracting from the actual activity. Since the subjects in the group with achievements interact longer with the Alexa skill but do not automatically process more vocabulary, such a conflict is possible in general and also in the experiment. As an implication, we recommend investigating how much additional time is needed for the presentation of the game design elements in the design process, and if it interferes with the motivation of the individuals.

Another aspect that may explain why the achievements do not contribute to the processing of more vocabulary can be attributed to their design in the study. Users could access their achievements, but there was no way to make them available to other users. What makes an achievement desirable is if it is available to a larger public and whether or not it is seen by others (Zichermann & Cunningham, 2011). It is more challenging to motivate users to collect achievements without comparison. Future studies should investigate if and how it is possible to access the game design element achievements for others.

Another design aspect that may have worked against the desired goal and thus weakened the effect of the achievements is the use of achievements for negative events. The goal in designing the achievements was to keep the subjects interested by offering a wide variety of unlocking options. Unlike the rest, two achievements, “every beginning is difficult” and “that is not what I meant, but okay” were unlocked when an answer was not quite appropriate or completely wrong. By awarding such negative achievements, the user is explicitly made aware of his mistake, which can have a demotivating effect (Kapp, 2012).

Perhaps achievements (and badges) are not suitable for a purely audio-based implementation. Since we based the selection of game design elements on those most intensively considered in gamification research (Christy & Fox, 2014;

Sailer & Sailer, 2021; Tan & Hew, 2016), which are usually implemented visually, it is not unlikely that these elements will not work in a purely audio-based environment (Friberg & Gärdenfors, 2004). To identify more appropriate game design elements for audio-gamification, an analysis of audio games would be helpful. These elements could then be applied to a similar context as in our study.

As for the second hypothesis, a statistically significant effect of both game design elements on learning outcomes was obtained. This result also supports the findings of previous research on gamification in education (Denny et al., 2018; Sailer & Homner, 2020). Thus, audio-gamification offers the potential to enhance further positive effects on learning outcomes through IVAs, such as those obtained by Dizon (2020).

Contrary to the results of Dizon (2020), and Hsu et al. (2021), a positive effect on the improvement of listening comprehension was found in the study. The different conditions of the studies provide a possible explanation for this. In both experiments, Dizon (2020) and Hsu et al. (2021) subjects trained their listening comprehension by interacting with an IVA whose system language was set to the foreign language being learned. In our study, the basic integration with the IVA occurred in the user’s preselected language, which can be assumed to be the native language. Only the vocabulary words to be learned, which were asked one by one, had to be understood correctly. Dizon (2020) assumes that the lack of effect is possibly due to the excessively high speaking rate and the too difficult vocabulary of the IVA. Both factors could be reduced in the study setting.

The evaluation of the interaction time, which was carried out in addition to the hypotheses, also differed statistically between the groups. Thus, the results could support the conclusions of previous studies on the effect of gamification (Landers & Landers, 2014; Majuri et al., 2018; Monterrat et al., 2015). Furthermore, we confirm that this potential can also be transferred to the context of IVAs and that the positive effect can also be generated by purely audio-based game design elements. As already proven in a laboratory experiment before (Bräuer & Mazarakis, 2022), audio-gamification can increase user engagement with an application. We were able to reproduce these results in a real-life situation. Unlike Bräuer and Mazarakis (2022), where several game design elements were used at once, this study goes into detail about the effect of the two individual game design elements examined, i.e., achievements and leaderboards. Our results take on the findings of the earlier study further by showing that audio-gamification works, but not every successful gamification concept can be readily transferred to the IVA context. However, the significance of these results should be interpreted with caution since the dialogs in the gamified groups are significantly longer than in the control group. Nevertheless, it is interesting that the subjects were not tempted by Alexa’s more complex instructions to work on fewer vocabulary items, e.g., to always stop after 5 min regardless of which group they were in.

Despite the positive results of the study, when integrating audio-gamification into applications for IVAs, it should be

noted that its integration increases the complexity of the application. Luger and Sellen (2016) point out that to interact with an IVA, users need time to grasp the scope of interaction options. Unlike visual interfaces, which allow users to see numerous options simultaneously, an IVA, similar to an audio game, requires each option to be introduced by speech (Friberg & Gärdenfors, 2004). For example, the category selection was always read aloud completely, which would not be necessary for a similar program on a device with a screen. Because the incorporation of game design elements might enhance the complexity of the conversational user interface, the time required for users to become acquainted with an application is likely to grow. As a result, it should constantly be verified whether gamifying an application makes it more difficult to use.

In this study, users could receive an overview of their achievements or their current position on the leaderboard in the main menu. However, these features were only used three times by the subjects. On the one hand, this very low interaction with the overview-features could be interpreted as very little engagement with the gamification system. Nevertheless, this is not the case because even without actively calling up these features, the users were constantly made aware of their progress through the integrated game design elements. For successful gamified systems, the user-system interaction plays a crucial role (Liu et al., 2017). User-system interactions include system-user communications (Liu et al., 2017, p. 1015) and aspects which provide feedback and leaderboards to monitor progress (Liu et al., 2017, p. 1023). This is the case for our experiment. As the positive effect of the leaderboard condition shows, calling up the overview does not seem necessary to increase motivation. From a design perspective, it should be reconsidered whether the integration of such an overview is necessary or makes the system unnecessarily complex.

The dropout rate of 47% of the subjects who accessed the experiment without processing even one vocabulary is worth mentioning. These individuals probably accessed the Alexa skill by mistake or lost interest in the Alexa skill directly after the explanation at the beginning, e.g., when the statement came that data is recorded for research purposes. Privacy concerns could potentially be a reason for direct termination in this scenario. Because many users are suspicious about an IVA processing personal data (Lau et al., 2018; Liao et al., 2019), the explanation about data gathering for research objectives when starting the Alexa skill may dissuade these users. However, from the standpoint of research ethics, such an explanation should not be missing. In addition, also other studies report high dropout rates, e.g., 38% in an online experiment about visual preferences (Jun et al., 2017). So, although we had anticipated a lower rate, this relatively high rate is not a cause for concern. Still, 230 subjects could be analyzed inferentially to test the hypotheses.

In addition to the high rate of users who did not use the Alexa skill at all, it should also be noted that many only used the Alexa skill once. Overall, 37 subjects, or 16% of the users in the sample for analysis, used the Alexa skill more than once. Nevertheless, although this number seems low

initially, it is relatively high in comparison. This is because the recurrence rate of applications for IVAs is generally very low and currently lies at 6% on average (Planner, 2018). In comparison, the Alexa skill developed for this study was thus activated more frequently, but the audio-gamification still does not seem to have really motivated regular use.

However, to learn a foreign language with the IVA in the long run, it would be necessary to use the Alexa skill over a longer period of time. One reason for the low multiple uses of the Alexa skill could be the relatively low mean percentage of correctly answered questions of 39%. The application might have been designed too difficult. The Alexa skill lacked a sequence in which the vocabulary was translated once before it had to be answered for the first time. To promote learning success, follow-up studies should first include a practice sequence in which all vocabulary words at the current level are presented once with the appropriate translation.

Also, to the design challenge of implementing audio-gamification, providing a foreign language learning Alexa skill brings some pitfalls that should not go unmentioned. When pre-testing the Alexa skill, we sort out as many words as possible that the IVA could potentially misunderstand. Nevertheless, individual inputs may not be correctly understood. Such a misunderstanding could certainly harm the user's motivation. In the field experiment, we did not store the user's input, but only recorded *via* a counter how often a vocabulary item was classified as correct or incorrect. To be able to exclude the influence of such errors on the user's motivation in future studies, a manipulation check in the form of precise storage of the input should be performed.

The goal of current gamification research should be to demonstrate the effect of individual game design elements so that specific conclusions can be drawn (Govender & Arnedo-Moreno, 2021; Landers, 2014; Mazarakis, 2021; Mekler et al., 2017). The game design elements achievements and leaderboard were considered in two different experimental conditions in the study. However, both featured several game design elements (points, levels, and challenges). Thus, each element's effect could have been strengthened or weakened by combining elements. The findings allow general assumptions about the effects of audio-gamification, but a detailed analysis of the effects of achievements and leaderboards is only possible to a limited extent.

As mentioned earlier, user fear of personal data processing by an IVA is a challenge (Lau et al., 2018; Liao et al., 2019). Therefore, to avoid discouraging users, the study refrained from asking for demographic data about the subjects, such as age and gender. However, previous research has shown that factors, such as age and gender can influence the effect of gamification (Codish & Ravid, 2017; Jent & Janneck, 2018). Similarly, users' age or gender can influence their IVA usage behavior (Zellou et al., 2021). Therefore, demographic data are vital for contextualizing the results, which might occasionally yield different interpretations than initially believed. In addition, the language level of the subjects, which was not asked in this study, can be necessary for data interpretation. For example, it is possible that some

demographic groups, such as women or persons with better language skills, were simply overrepresented in particular groups, which could potentially explain why these groups differed significantly. Additional information about subjects would be helpful in subsequent studies to uncover such effects, although users may be deterred by these questions for privacy reasons. It should be noted, however, that there is no evidence in the field experiment that the biases described here may be present.

As a study by Leung et al. (2022) showed, the effect of gamification is not only influenced by the selection of different game design elements. Adaptation to individual characteristics is also relevant, where different game design elements can cause different effects on online learning. One aspect that can influence the effect of gamified learning systems, according to Leung et al. (2022), is the individual goal orientation of the learners. Therefore, they recommend applying personalized approaches to gamification in the field of education in the future. To optimally design the effect of audio-gamification in the field of vocabulary learning, it would therefore also be worth considering a personalized approach to gamification for future studies, which is already acknowledged as an important topic for gamification in general (Mazarakis, 2021).

Another limitation of the study is that it did not record what type of device the users were using. Dizon et al. (2022) found that in a study with an Echo Show, unlike previous studies with a speaker without a display, new problems arose during the first interaction with the device. For example, users attempted to type their responses *via* the display instead of giving a verbal response. Even though in a field experiment, it is highly likely that users already have experience with IVAs, especially the time users spend with the Alexa skill could be influenced by such issues. To counteract this, care was taken during implementation not to show any information on the display. Only the name of the Alexa skill was displayed.

Although SDT is often used only to distinguish between intrinsic and extrinsic motivation (Koivisto & Hamari, 2019), SDT offers a much broader range of explanatory possibilities. For example, competitive game design elements may be perceived as introjected regulation of extrinsic motivation and have a favorable short-term influence in various fields. This complexity is already acknowledged (Tyack & Mekler, 2020), and future studies should use the full potential of the self-determination theory.

As Ruan et al. (2019) already showed, learning vocabulary *via* voice-based systems is time-consuming but, at the same time, seems to be more motivating than classical learning with article. The gamification integration, as done in our study, definitely increases the interaction time. Speech was mainly used, using audio feedback only when a challenge was completed or an achievement was acquired. However, when looking at audio games, music and sound are common to represent the mechanics of the game (Ciccio & Quesada, 2018). So, for example, more diverse sounds or melodies could be used after unlocking an achievement. Similarly, the oral feedback that users receive when

answering vocabulary could be replaced by audio feedback, as is often found in apps and games (Thiebes et al., 2014). In the field of audio game development, Friberg and Gärdenfors (2004) recommend providing individual sounds with several layers of information. For example, a sound can be used to indicate whether a machine is running or switched off and simultaneously convey how far away it is from the player. One recommendation that can be derived from our experience is to look for ways to integrate game design elements so that they can also be presented in parallel to the actual application. In this way, it might be possible to take advantage of gamification in the context of IVAs without losing the user's attention by making the interaction time too long.

6. Conclusion

In this long-term field experiment, quantitative data on the usage behavior of a gamified Amazon Alexa skill was analyzed. The research question is answered how the combination of IVAs and audio-gamification in the field of foreign language learning influences motivation and learning success. Using a field experiment instead of a lab experiment provides robust quantitative data in the area of conversational interfaces, in contrast to the primarily conducted qualitative studies. Data from 230 subjects were analyzed for one year. This revealed insights into design approaches and their effects on subject engagement and learning success. Thus, building on this study, the first conclusions can be drawn about which game design elements are particularly suitable for learning a new language in the context of voice assistants, e.g., leaderboards. This study provides results that contribute to opening new design perspectives for IVA research, as well as highlighting the aspect of audio elements from the perspective of gamification research.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Paula Bräuer  <http://orcid.org/0000-0001-5903-8829>
Athanasios Mazarakis  <http://orcid.org/0000-0001-9943-0382>

References

- Adams, S. (2019, August 31). *Game of tongues: How duolingo built a \$700 million business with its addictive language-learning app*. Forbes. <https://www.forbes.com/sites/susanadams/2019/07/16/game-of-tongues-how-duolingo-built-a-700-million-business-with-its-addictive-language-learning-app/>
- Antin, J., & Churchill, E. F. (2011). Badges in social media: A social psychological perspective. In *CHI 2011 Gamification Workshop Proceedings* (pp. 1–4).
- Babic, S., Orehovacki, T., & Etinger, D. (2018). Perceived user experience and performance of intelligent personal assistants employed in higher education settings. In *2018 41st International Convention on Information and Communication Technology, Electronics and*

- Microelectronics (MIPRO)* (pp. 830–834). <https://doi.org/10.23919/MIPRO.2018.8400153>
- Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, 30, 100322. <https://doi.org/10.1016/j.edurev.2020.100322>
- Benner, D., Schöbel, S., Süess, C., Baechle, V., Janson, A. (2022). Level-up your learning – Introducing a framework for gamified educational conversational agents. In *Wirtschaftsinformatik 2022 Proceedings*. <https://aisel.aisnet.org/wi2022/hci/hci/5>
- Bouchrika, I., Harrati, N., Wanick, V., & Wills, G. (2021). Exploring the impact of gamification on student engagement and involvement with E-learning systems. *Interactive Learning Environments*, 29(8), 1244–1257. <https://doi.org/10.1080/10494820.2019.1623267>
- Bräuer, P., & Mazarakis, A. (2019). Badges or a leaderboard? How to gamify an augmented reality warehouse setting. In J. Koivisto & Hamari (Eds.), *Proceedings of the 3rd International GamiFIN Conference—GamiFIN 2019* (Vol. 2359, pp. 229–240). CEUR-WS.org.
- Bräuer, P., & Mazarakis, A. (2022). “Alexa, can we design gamification without a screen?”—Implementing cooperative and competitive audio-gamification for intelligent virtual assistants. *Computers in Human Behavior*, 135, 107362. <https://doi.org/10.1016/j.chb.2022.107362>
- Cermak-Sassenrath, D. (2019). Current challenges in gamification identified in empirical studies. In R. Ørngreen, M. Buhl, & B. Meyer (Eds.), *Proceedings of the 18th European Conference on e-Learning (ECEL)* (pp. 119–127). Academic Conferences and Publishing International Limited.
- Change the language of Google Assistant—Android—Google Nest Help* (2021). <https://support.google.com/googlenest/answer/7550584?hl=en&co=GENIE.Platform%3DAndroid&zippy=%2Cgoogle-nest-mini-gen>
- Chen, H. H.-J., Yang, C. T.-Y., & Lai, K. K.-W. (2020). Investigating college EFL learners’ perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments*, 1–16. <https://doi.org/10.1080/10494820.2020.1833043>
- Christy, K. R., & Fox, J. (2014). Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women’s math performance. *Computers & Education*, 78, 66–77. <https://doi.org/10.1016/j.compedu.2014.05.005>
- Ciccio, J. A., & Quesada, L. (2018). Framework for creating audio games for intelligent personal assistants. In T. Ahram & C. Falcão (Eds.), *Advances in human factors in wearable technologies and game design* (pp. 204–214). Springer International Publishing. https://doi.org/10.1007/978-3-319-60639-2_21
- Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., Aylett, M., Cabral, J., Munteanu, C., Edwards, J., & R Cowan, B. (2019). The state of speech in HCI: Trends, themes and challenges. *Interacting with Computers*, 31(4), 349–371. <https://doi.org/10.1093/iwc/iwz016>
- Codish, D., & Ravid, G. (2017). Gender moderation in gamification: Does one size fit all? In *Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2017.244>
- Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D., & Bandeira, N. (2017). “What can I help you with?”: Infrequent users’ experiences of intelligent personal assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1–12). <https://doi.org/10.1145/3098279.3098539>
- de Barcelos Silva, A., Gomes, M. M., da Costa, C. A., da Rosa Righi, R., Barbosa, J. L. V., Pessin, G., De Doncker, G., & Federizzi, G. (2020). Intelligent personal assistants: A systematic literature review. *Expert Systems with Applications*, 147, 113193. <https://doi.org/10.1016/j.eswa.2020.113193>
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talae, E., & Noroozi, O. (2021). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning*, 34(7), 934–957. <https://doi.org/10.1080/09588221.2019.1648298>
- Denny, P., McDonald, F., Empson, R., Kelly, P., & Petersen, A. (2018). Empirical support for a causal relationship between gamification and learning outcomes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). <https://doi.org/10.1145/3173574.3173885>
- 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). <https://doi.org/10.1145/2181037.2181040>
- Develop Skills in Multiple Languages | Alexa Skills Kit* (2021). Amazon (Alexa). <https://developer.amazon.com/de-DE/docs/alexa/custom-skills/develop-skills-in-multiple-languages.html>
- Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A critical review. *International Journal of Educational Technology in Higher Education*, 14(1), 9. <https://doi.org/10.1186/s41239-017-0042-5>
- Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Educational Technology & Society*, 18(3), 75–88.
- Dizon, G. (2017). Using intelligent personal assistants for second language learning: A case study of Alexa. *TESOL Journal*, 8(4), 811–830. <https://doi.org/10.1002/tesj.353>
- Dizon, G. (2020). Evaluating intelligent personal assistants for L2 listening and speaking development. *Language Learning & Technology*, 24(1), 16–26. <http://hdl.handle.net/10125/44705>
- Dizon, G. (2021). Affordances and constraints of intelligent personal assistants for second-language learning. *RELC Journal*, 003368822110205. <https://doi.org/10.1177/00336882211020548>
- Dizon, G., Tang, D., & Yamamoto, Y. (2022). A case study of using Alexa for out-of-class, self-directed Japanese language learning. *Computers and Education: Artificial Intelligence*, 3, 100088. <https://doi.org/10.1016/j.caeai.2022.100088>
- Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- Friberg, J., & Gärdenfors, D. (2004). Audio games: New perspectives on game audio. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology – ACE’04* (pp. 148–154). <https://doi.org/10.1145/1067343.1067361>
- Garcia, F. E., de Almeida Neris, V. P., Garcia, F. E., & de Almeida Neris, V. P. (2013). Design guidelines for audio games. In M. Kurosu & M. Kurosu (Eds.), *Proceedings of the 15th International Conference on Human-Computer Interaction* (Vol. 8005, pp. 229–238). Springer. https://doi.org/10.1007/978-3-642-39262-7_26
- Govender, T., & Arnedo-Moreno, J. (2021). An analysis of game design elements used in digital game-based language learning. *Sustainability*, 13(12), 6679. <https://doi.org/10.3390/su13126679>
- Groening, C., & Binnewies, C. (2019). “Achievement unlocked!”—The impact of digital achievements as a gamification element on motivation and performance. *Computers in Human Behavior*, 97, 151–166. <https://doi.org/10.1016/j.chb.2019.02.026>
- Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? – A literature review of empirical studies on gamification. In *47th Hawaii International Conference on System Sciences (HICSS)* (pp. 3025–3034). <https://doi.org/10.1109/HICSS.2014.377>
- Hsu, H.-L., Chen, H. H.-J., & Todd, A. G. (2021). Investigating the impact of the Amazon Alexa on the development of L2 listening and speaking skills. *Interactive Learning Environments*, 1–14. <https://doi.org/10.1080/10494820.2021.2016864>
- Huang, R., Ritzhaupt, A. D., Sommer, M., Zhu, J., Stephen, A., Valle, N., Hampton, J., & Li, J. (2020). The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology Research and Development*, 68(4), 1875–1901. <https://doi.org/10.1007/s11423-020-09807-z>

- IOS and iPadOS-Feature Availability (2021). Apple. <https://www.apple.com/ios/feature-availability/>
- Istrate, A. M. (2019). *The impact of the virtual assistant (VA) on language classes* (1.0, 163840. B) [Data set]. ADLRO. <https://doi.org/10.12753/2066-026X-19-040>
- Jent, S., & Janneck, M. (2018). Using gamification to enhance user motivation: The influence of gender and age. In L. E. Freund & W. Cellary (Eds.), *Advances in the human side of service engineering* (Vol. 601, pp. 3–10). Springer International Publishing. https://doi.org/10.1007/978-3-319-60486-2_1
- Jun, E., Hsieh, G., & Reinecke, K. (2017). Types of motivation affect study selection, attention, and dropouts in online experiments. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 1–15. <https://doi.org/10.1145/3134691>
- Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. John Wiley & Sons.
- Kinsella, B. (2019, March 7). *U.S. smart display user base grew 558% in 2018 and more than doubled in second half of the year, Amazon holds two-thirds market share*. Voicebot.Ai. <https://voicebot.ai/2019/03/07/u-s-smart-display-user-base-grew-558-in-2018-and-more-than-doubled-in-second-half-of-the-year-amazon-holds-two-thirds-market-share/>
- Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, 45, 191–210. <https://doi.org/10.1016/j.ijinfomgt.2018.10.013>
- La Guardia, J. G., Ryan, R. M., Couchman, C. E., & Deci, E. L. (2000). Within-person variation in security of attachment: A self-determination theory perspective on attachment, need fulfillment, and well-being. *Journal of Personality and Social Psychology*, 79(3), 367–384. <https://doi.org/10.1037/0022-3514.79.3.367>
- Landers, R. N. (2014). Developing a theory of gamified learning: linking serious games and gamification of learning. *Simulation & Gaming*, 45(6), 752–768. <https://doi.org/10.1177/1046878114563660>
- 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. <https://doi.org/10.1177/1046878114563662>
- Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, are you listening?: Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–31. <https://doi.org/10.1145/3274371>
- Legaki, N.-Z., Xi, N., Hamari, J., Karpouzis, K., & Assimakopoulos, V. (2020). The effect of challenge-based gamification on learning: An experiment in the context of statistics education. *International Journal of Human-Computer Studies*, 144, 102496. <https://doi.org/10.1016/j.ijhcs.2020.102496>
- Leung, A. C. M., Santhanam, R., Kwok, R. C.-W., & Yue, W. T. (2022). Could gamification designs enhance online learning through personalization? Lessons from a field experiment. *Information Systems Research*, 1–23. <https://doi.org/10.1287/isre.2022.1123>
- Liao, Y., Vitak, J., Kumar, P., Zimmer, M., & Kritikos, K. (2019). Understanding the role of privacy and trust in intelligent personal assistant adoption. In N. G. Taylor, C. Christian-Lamb, M. H. Martin, & B. Nardi (Eds.), *Information in contemporary society* (Vol. 11420, pp. 102–113). Springer International Publishing. https://doi.org/10.1007/978-3-030-15742-5_9
- Liu, D., Santhanam, R., & Webster, J. (2017). Toward meaningful engagement: A framework for design and research of gamified information systems. *MIS Quarterly*, 41(4), 1011–1034. <https://doi.org/10.25300/MISQ/2017/41.4.01>
- Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., & Maher, C. (2017). Does gamification increase engagement with online programs? A systematic review. *PLOS One*, 12(3), e0173403. <https://doi.org/10.1371/journal.pone.0173403>
- Lopatovska, I., Griffin, A. L., Gallagher, K., Ballingall, C., Rock, C., & Velazquez, M. (2020). User recommendations for intelligent personal assistants. *Journal of Librarianship and Information Science*, 52(2), 577–591. <https://doi.org/10.1177/0961000619841107>
- Luger, E., & Sellen, A. (2016). “Like having a really bad PA”: The gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5286–5297). <https://doi.org/10.1145/2858036.2858288>
- Maeda, J. (n.d.). *Frequency level checker*. Retrieved January 16, 2022, from <https://chuta.cegloc.tsukuba.ac.jp/flc/index.html>
- Majuri, J., Koivisto, J., Hamari, J. (2018). Gamification of education and learning: A review of empirical literature. In *Proceedings of the 2nd International GamiFIN Conference, GamiFIN 2018* (pp. 11–19).
- Marczewski, A. (2018). *Even Ninja monkeys like to play: Gemification, gaze thinking and motivational design* (Unicorn ed.).
- Mazarakis, A. (2015). Using gamification for technology enhanced learning: The case of feedback mechanisms. *Bulletin of the IEEE Technical Committee on Learning Technology*, 4(17), 6–9. https://tc.computer.org/tclt/wp-content/uploads/sites/5/2018/01/2_Mazarakis.pdf
- Mazarakis, A. (2021). Gamification reloaded: Current and future trends in gamification science. *i-com*, 20(3), 279–294. <https://doi.org/10.1515/icom-2021-0025>
- Mazarakis, A., & Bräuer, P. (2018). Gamification is working, but which one exactly? Results from an experiment with four game design elements. In *Proceedings of the Technology, Mind, and Society* (Vol. 22, p. 1). <https://doi.org/10.1145/3183654.3183667>
- Mazarakis, A., & Bräuer, P. (2022). Gamification is working, but which one exactly? Results from an experiment with four game design elements. *International Journal of Human-Computer Interaction*, 1–16. <https://doi.org/10.1080/10447318.2022.2041909>
- McTear, M., Callejas, Z., & Griol, D. (2016). *The conversational interface*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-32967-3>
- Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, 71, 525–534. <https://doi.org/10.1016/j.chb.2015.08.048>
- Monterrat, B., Desmarais, M., Lavoué, É., & George, S. (2015). A player model for adaptive gamification in learning environments. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Artificial intelligence in education* (Vol. 9112, pp. 297–306). Springer International Publishing. https://doi.org/10.1007/978-3-319-19773-9_30
- Moussalli, S., & Cardoso, W. (2020). Intelligent personal assistants: Can they understand and be understood by accented L2 learners? *Computer Assisted Language Learning*, 33(8), 865–890. <https://doi.org/10.1080/09588221.2019.1595664>
- Murad, C., Munteanu, C., Clark, L., & Cowan, B. R. (2018). Design guidelines for hands-free speech interaction. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct – MobileHCI’18* (pp. 269–276). <https://doi.org/10.1145/3236112.3236149>
- Namaghi, S. A. O., Safae, S. E., & Sobhanifar, A. (2015). The effect of shyness on English speaking scores of Iranian EFL learners. *Journal of Literature, Language and Linguistics*, 12, 22–28.
- Nobles, A. L., Leas, E. C., Caputi, T. L., Zhu, S.-H., Strathdee, S. A., & Ayers, J. W. (2020). Responses to addiction help-seeking from Alexa, Siri, Google Assistant, Cortana, and Bixby intelligent virtual assistants. *NPJ Digital Medicine*, 3(1), 11. <https://doi.org/10.1038/s41746-019-0215-9>
- Ortiz-Rojas, M., Chiluiza, K., & Valcke, M. (2019). Gamification through leaderboards: An empirical study in engineering education. *Computer Applications in Engineering Education*, 27(4), 777–788. <https://doi.org/10.1002/cae.12116>
- Planner, A. (2018, July 5). *How many users should you expect for your voice skill?* <https://www.screenmedia.co.uk/news/how-many-users-should-you-expect-for-your-voice-skill/>
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Pyae, A., & Scifleet, P. (2018). Investigating differences between native English and non-native English speakers in interacting with a voice

- user interface: A case of Google Home. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (pp. 548–553). <https://doi.org/10.1145/3292147.3292236>
- Pyae, A., & Scifleet, P. (2019). Investigating the role of user's English language proficiency in using a voice user interface: A case of Google Home smart speaker. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–6). <https://doi.org/10.1145/3290607.3313038>
- Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, L., & Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons*, 58(4), 411–420. <https://doi.org/10.1016/j.bushor.2015.03.006>
- Ruan, S., Jiang, L., Xu, J., Tham, B. J.-K., Qiu, Z., Zhu, Y., Murnane, E. L., Brunskill, E., & Landay, J. A. (2019). QuizBot: A dialogue-based adaptive learning system for factual knowledge. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). <https://doi.org/10.1145/3290605.3300587>
- Ruan, S., Jiang, L., Xu, Q., Liu, Z., Davis, G. M., Brunskill, E., & Landay, J. A. (2021). EnglishBot: An AI-powered conversational system for second language learning. In *26th International Conference on Intelligent User Interfaces* (pp. 434–444). <https://doi.org/10.1145/3397481.3450648>
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54–67. <https://doi.org/10.1006/ceps.1999.1020>
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.
- Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32(1), 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Sailer, M., & Sailer, M. (2021). Gamification of in-class activities in flipped classroom lectures. *British Journal of Educational Technology*, 52(1), 75–90. <https://doi.org/10.1111/bjet.12948>
- 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. <https://doi.org/10.1016/j.chb.2016.12.033>
- Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74, 14–31. <https://doi.org/10.1016/j.ijhcs.2014.09.006>
- Skidmore, L., & Moore, R. K. (2019). Using Alexa for flashcard-based learning. In *Interspeech 2019* (pp. 1846–1850). <https://doi.org/10.21437/Interspeech.2019-2893>
- Speech Synthesis Markup Language (SSML) Reference | Alexa Skills Kit* (n.d.). Amazon (Alexa). Retrieved July 17, 2021, from <https://developer.amazon.com/en-US/docs/alexa/custom-skills/speech-synthesis-markup-language-ssml-reference.html>
- Tan, M., & Hew, K. F. (2016). Incorporating meaningful gamification in a blended learning research methods class: Examining student learning, engagement, and affective outcomes. *Australasian Journal of Educational Technology*, 32(5), 19–34. <https://doi.org/10.14742/ajet.2232>
- Tejedor-Garcia, C., Escudero-Mancebo, D., Cardenoso-Payo, V., & Gonzalez-Ferreras, C. (2020). Using challenges to enhance a learning game for pronunciation training of English as a second language. *IEEE Access*, 8, 74250–74266. <https://doi.org/10.1109/ACCESS.2020.2988406>
- Tejedor-Garcia, C., Escudero-Mancebo, D., Gonzalez-Ferreras, C., Cámara-Arenas, E., & Cardenoso-Payo, V. (2016). Improving L2 production with a gamified computer-assisted pronunciation training tool, Tiptoptalk! In *Proceedings of IberSPEECH* (pp. 177–186).
- Thiebes, S., Lins, S., & Basten, D. (2014). Gamifying information systems—A synthesis of gamification mechanics and dynamics. In *ECIS 2014 Proceedings* (pp. 1–17).
- Tyack, A., & Mekler, E. D. (2020). Self-determination theory in HCI games research: Current uses and open questions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–22). <https://doi.org/10.1145/3313831.3376723>
- Vansteenkiste, M., & Ryan, R. M. (2013). On psychological growth and vulnerability: Basic psychological need satisfaction and need frustration as a unifying principle. *Journal of Psychotherapy Integration*, 23(3), 263–280. <https://doi.org/10.1037/a0032359>
- Vitkauskaitė, E., & Gatautis, R. (2018). Points for posts and badges to brand advocates: The role of gamification in consumer brand engagement. In *Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2018.143>
- Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Wharton Digital Press.
- Wu, Y., Rough, D., Bleakley, A., Edwards, J., Cooney, O., Doyle, P. R., Clark, L., & Cowan, B. R. (2020). See what I'm saying? Comparing intelligent personal assistant use for native and non-native language speakers. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1–9). <https://doi.org/10.1145/3379503.3403563>
- Young, H. (2014, November 7). Do young people care about learning foreign languages? *The Guardian*. <https://www.theguardian.com/education/2014/nov/07/-sp-do-young-people-care-about-learning-for-foreign-languages-data>
- Zellou, G., Cohn, M., & Ferenc Segedin, B. (2021). Age- and gender-related differences in speech alignment toward humans and voice-AI. *Frontiers in Communication*, 5, 600361. <https://doi.org/10.3389/fcomm.2020.600361>
- Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. O'Reilly Media.

About the authors

Paula Bräuer is a research assistant and PhD candidate at Kiel University since 2018. She has a master's degree in business informatics and researches gamification in various contexts, such as augmented reality and intelligent virtual assistants.

Athanasios Mazarakis is a post-doc for Web Science at ZBW–Leibniz Information Centre for Economics and has been working on gamification and incentives in the interdisciplinary field between computer science, economics, and psychology for over a decade. Numerous publications on gamification and successful workshop organizations complete his competence profile.

CHAPTER

SIX

Game Design Elements from Alexa Games

The study presented in this chapter has been published as a full paper at the GamifFIN Conference 2023 and has received a double-blind peer review. The results were presented at the GamiFIN conference in Levi in April 2023, where the paper was awarded the best paper.

Bräuer, P., Werner, F., & Mazarakis, A. (2023). Extracting Game Design Elements from Voice-Enabled Games: A Review of Amazon Alexa Skills. *Proceedings of the 7th International GamiFIN Conference*, 47–56.

The data were collected in the context of a seminar in 2022. The seminar was led by the author and Dr. Athanasios Mazarakis, who also jointly interrogated the design of the study. The data collection was carried out by the student participants who attended the seminar. The analysis of the data and manuscript preparation were taken by the author and Florian Werner and revised by Dr. Athanasios Mazarakis before submission.

6.1. Preface to Study 4

Studies two and three show that audio-gamification can motivate individuals to use IVAs. However, the effects of different game design elements differ, as shown in studies one and three. Classical game design elements such as badges, whose effects have been widely demonstrated in studies (Hakulinen et al., 2015; Hamari, 2017; Mazarakis & Bräuer, 2020), are challenging to implement without a screen, and no effect on motivation could be achieved with achievements, which are supposed to be delivered via voice (Bräuer & Mazarakis, 2022b). It seems that audio-gamification should use different game design elements than those that have proven successful so far. This thesis's fourth and

final study addresses the question of what these elements might be. To answer this question, we systematically analyzed which game design elements can be found in the games offered in the Amazon Alexa store. For the analysis, the gamification framework of Werbach and Hunter (2012) was used to search for known game design elements in the Alexa skills. This last study shows the possibilities for further research on this topic.

6.2. Study 4 (publication)

Extracting Game Design Elements from Voice-Enabled Games: A Review of Amazon Alexa Skills

Paula Bräuer¹, Florian Werner¹ and Athanasios Mazarakis²

¹ Kiel University, Christian-Albrechts-Platz 4, Kiel, 24118, Germany

² ZBW-Leibniz Information Centre for Economics, Düsternbrooker Weg 120, Kiel, 24105, Germany

Abstract

Intelligent virtual assistants such as Amazon Alexa, Siri, and Google Assistant are mainly used to request information or play music. However, they can also be used to play voice-enabled games. So far, very little research has looked at this new form of gaming. To fill this research gap, 29 games from the Amazon Alexa Store were systematically studied and compared with the three-category-model by Werbach and Hunter, identifying which game design elements are used and how frequently. This allowed us to exemplify the differences between the game design elements used in voice-enabled games and those frequently used in video games and gamification. Furthermore, two game categories were identifiable that were correlated with the game design elements. The investigation of the Alexa games can help to draw new conclusions for the further development of voice-enabled games and to provide implications for the gamification of speech-based applications.

Keywords

Voice Assistant, Intelligent Personal Assistant, Voice Interaction, Gamification, Game Design

1. Introduction

In recent years, intelligent virtual assistants (IVAs) such as Amazon Alexa and Google Assistant have grown in importance in research and practice [4]. IVAs enable new forms of human-machine interaction through spoken language [24]. This form of interaction is more intuitive than the usual mouse and keyboard input [26] and offers advantages such as facilitating multitasking, e.g., when cooking or driving, or improving accessibility for people with limited vision. Voice control is also being used as a new element of interaction in video games, although this presents some challenges [2]. Similar to apps for smartphones, the functionality of an IVA can be enhanced by applications from external developers. These applications for IVAs are often referred to as "skills" [18]. As IVAs continue to improve, new skills are being added to their scope

that goes beyond obtaining weather information or playing music [7].

Meanwhile, more and more games for IVAs are being developed. The category "games and quizzes" now accounts for around 15% of the skills in the German Alexa Store and is the second largest of the 21 categories after "music and audio." Despite this, there is currently minimal research on games for IVAs found in the literature [4].

Unlike video games, most voice-controlled games do not use visual elements. As a result, all information is conveyed via speech, music, or other sounds [28]. This is partly because IVAs are mostly used via smart speakers, many of which, like the Amazon Echo Dot or the Google Nest Mini, do not have a screen. The current crisis at Amazon, which is struggling with major losses in the Alexa division [6], indicates how challenging it is to make skills appealing. One way to make skill design more attractive and to motivate more

7th International GamiFIN Conference 2023 (GamiFIN 2023), April 18-21, 2023, Lapland, Finland.

EMAIL: p.braeuer@zbw.eu (A. 1); a.mazarakis@zbw.eu (A. 3)

ORCID: 0000-0001-5903-8829 (A. 1); 0000-0001-9943-0382 (A. 3)



© 2020 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

extensive interactions with an IVA could be gamification [5,8,9,23].

Our study aims to systematically analyze voice-enabled games to gain insights into which game design elements are used in the context of IVAs. Through the analysis, based on the game design elements used, we hope to derive implications of which elements are suitable for the gamification of IVA applications. This leads us to the research question:

RQ: What are the dominant game design elements used in voice-enabled games for Amazon Alexa?

To the best of our knowledge, there are currently no comprehensive studies of Alexa game skills, therefore our study is pioneering the field. Chapter two provides some background information and related work. In the following chapter, we describe how the games were classified. The results are presented in the fourth chapter, and then a discussion follows on how they can be interpreted. Finally, limitations, future research, and a conclusion are presented. Building on these results, we hope to derive recommendations as to which game design elements are particularly well suited for the design of voice-based games and possibly also for the gamification of voice-based applications.

2. Related Work

In the literature, games for IVAs have received relatively little attention. A paper by Benner et al. [5] addresses the development of a framework for the gamification of educational conversational agents. For the framework, existing literature on chatbots and conversational agents was examined. These technologies also work with speech recognition but are based on written and not spoken language as IVAs. Accordingly, a screen is used to integrate classical visual game design elements, which is not possible with most smart speakers.

Cicció and Quesada [15] present a framework to assist in creating games for IVAs. The framework is based on the game development experiences of the authors and distinguishes four components: game design, audio game manager, natural language processor, and help modules. The last three components represent technical aspects that need to be considered to enable audio game control as well as speech input. The game design includes five elements: game mechanics, goals, story, rewards, and user progression. The

authors define game mechanics "as the way players interact and give input to the game" [15]. The framework provides a good starting point, but the set of proposed game design elements seems very limited compared to other frameworks from the gamification field [20,31,32].

To the best of our knowledge, no work has systematically analyzed which game design elements are used in games or gamified applications for IVAs. However, several papers have studied gamified apps for smartphones [14,30,33].

One example is the work of Schmidt-Kraepelin et al. [29], who focus on mHealth apps and investigate which gamification concepts can be found in different apps. In addition to capturing the game design elements found in the apps studied, the authors also look at the relationship between the number of game design elements and user ratings. For apps from the Apple Store, a positive correlation was found between user ratings and the number of game design elements.

Existing app reviews are either based on apps from the Google Play Store or Apple Store [29,33] or on existing literature [14,30]. Since no literature exists so far, an analysis of IVA skills is performed for our study. To systematically analyze which game design elements can already be found in game skills, an existing model will be used. Werbach and Hunter [32] describe a three-category model, which is one of the best-known gamification frameworks [17,27] with over 4000 citations on Google Scholar. The authors divide 30 elements in their model into three hierarchical categories, which are illustrated in the form of a pyramid. Dynamics represent the top level of the model and are described as the overall picture of a gamified application that cannot necessarily be directly influenced [32]. The mechanics are the model's intermediate level and are designed to depict basic operations in a gamified application [32]. Finally, the components serve as the model's foundation. They represent the specification of mechanics and dynamics they embody them and make them tangible [32]. An overview of all 30 elements and their allocation to the three categories of the model is shown in Table 1.

This model by Werbach and Hunter will be used to answer our research question and review IVA games because it is a common model used in gamification research [17,27]. Moreover, it provides a relatively wide range of systematically constructed elements, and each element is at least minimally described, which is essential for a consistent classification of the games.

3. Method

3.1. Selection of Skills

In our study, only skills from the Amazon Alexa Store were examined since Amazon Alexa is the most widely used smart speaker worldwide [10]. In addition, Google Assistant, the largest competitor of Amazon Alexa, only suggested around 20 applications in the "games and fun" category². In contrast, an examination of the "games and quizzes" category in the Amazon Alexa Store suggested more than five hundred times the number of skills³.

The selection of the Alexa skills to be studied was made on December 2021. For the analysis, the potential Alexa skills were narrowed down by first focusing on the German Alexa skill market. The German Alexa market was selected to prevent misunderstandings and problems with speech recognition, as the native language of the testers is German. By the time of data collection, the German Amazon Alexa Store provided more than 10,000 skills. Next, the selection of Alexa skills is focused on the "games and quizzes" category. This category consists of three subcategories: "games," "knowledge and quiz" and "game information and accessories." The last category does not contain any games and is therefore excluded. The remaining two categories contained 1,570 Alexa skills.

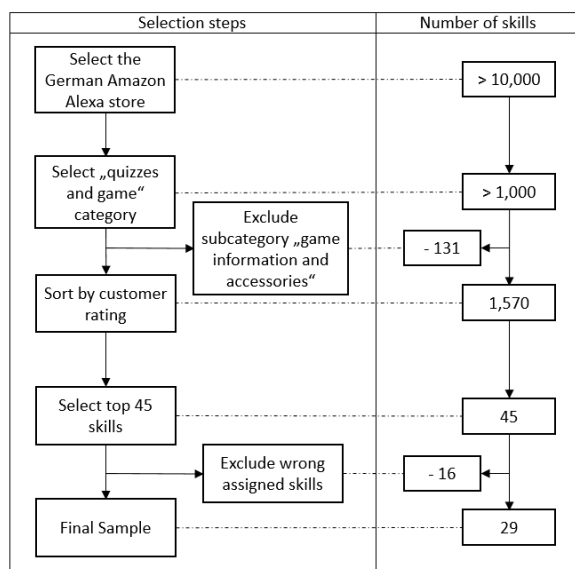


Figure 1: Flowchart of the Alexa skill selection

² <https://assistant.google.com/explore/c/5/?hl=de-DE>

These 1,570 Alexa skills were sorted in descending order by customer rating. This list was reduced to the top 45 rated games. The top 45 list still contained Alexa skills that were assigned to the wrong category by the developers. After reading the Alexa skill descriptions and testing the skills, these were also removed from our analysis. This reduced the number of Alexa skills to be reviewed to 29. The selection process for the Alexa skills for analysis depicts Figure 1. The list of Alexa skills with their respective user ratings and classification according to the Amazon Store can be found in the appendix.

3.2. Game design element classification

Next, each Alexa skill of the sample was classified regarding the model by Werbach and Hunter [32]. To reduce classification subjectivity, each Alexa skill on the list was played individually and independently on an Alexa Echo Dot 2 by two researchers. In addition, an informal script was created during testing. The plot of the game and special features are documented in this script. There were two rounds of classification: one while playing each game and a second one using the script. The researchers merely classified whether a given game design element was applied but not to what extent or quality it was implemented. After several days, each game was started again from the beginning. If no additional game design element was identified after another run, the game was not played again. This resulted in an evaluation table for each researcher that described which game contained which game design elements.

Inter-coder reliability was measured using Cohen's Kappa. A Cohen's Kappa coefficient of .64 was calculated overall. Thus, a substantial level of agreement between coders was achieved following established guidelines [22].

4. Results

All the 29 Alexa skills examined contained at least one of the game design elements according to Werbach and Hunter's model [32]. The overall mean value of the implemented game design elements was 6.97 per Alexa skill, with a standard

³ https://www.amazon.de/s?bbn=10068461031&rh=n%3A10068460031%2Cn%3A10536644031&dc&qid=1672997334&mid=10068461031&ref=lp_10068461031_nr_n_18

deviation of 2.92. A maximum of 16 and a minimum of three game design elements were identified in individual Alexa skills. Looking at the three categories of the model, on average 2.31 (SD 0.71) dynamics, 2.86 (SD 1.41) mechanics, and 1.79 (SD 1.86) components are found per Alexa skill. Table 1 summarizes the frequency at which the individual elements were discovered.

Table 1: List of elements according to Werbach and Hunter and the frequency at which they were found in the 29 Alexa games analyzed.

| Game Design Element | Number of Games |
|----------------------|-----------------|
| Dynamics | |
| Constraints | 25 |
| Emotions | 29 |
| Narrative | 13 |
| Progression | 1 |
| Relationships | 1 |
| Mechanics | |
| Challenges | 26 |
| Chance | 5 |
| Competition | 12 |
| Cooperation | 0 |
| Feedback | 18 |
| Resource Acquisition | 6 |
| Rewards | 4 |
| Transactions | 1 |
| Turns | 3 |
| Win States | 9 |
| Components | |
| Achievements | 0 |
| Avatars | 12 |
| Badges | 0 |
| Boss Fights | 1 |
| Collections | 0 |
| Combat | 1 |
| Content Unlocking | 1 |
| Gifting | 0 |
| Leaderboards | 3 |
| Levels | 1 |
| Points | 7 |
| Quests | 13 |
| Social Graphs | 3 |
| Teams | 1 |
| Virtual Goods | 9 |

Two types of games could be identified during the testing of the Alexa skills. Some games are characterized by the game design element narrative. Generally, the players are guided through a story and can make decisions at various points influencing the course of action. An example is the game "Heldentum für Anfänger"

(roughly translated as "heroism for beginners"), where the players can achieve eight different endings by their decisions.

The second type are quiz games. In the Amazon Alexa Store as well, the distinction between games and quizzes is made, suggesting that quizzes take on a special role in voice-enabled games. These quiz games consist of a varying number of questions that the user has to answer. The questions in such a quiz can vary daily, as in "Quiz des Tages", which could possibly motivate users to interact with the Alexa skill regularly. Thirteen of the Alexa Skills examined could be assigned to the "narrative game" category and another thirteen to the "quiz game" category. The three remaining skills from the sample cannot be assigned to either category but also do not form a recognizable category of their own. The list of Alexa skills classified into narrative and quiz games can be found in the Appendix.

A statistical analysis shows the correlations between the two types of games and the game design elements, according to Werbach and Hunter [32]. The data was initially checked for normal distribution. The Shapiro-Wilk test becomes statistically significant for all variables, so the data are not normally distributed, and a non-parametric test (Spearman correlation) is applied for further analyses. Since all game design elements were correlated with the two game types in this exploratory analysis, the alpha level was set to .001 using the Bonferroni correction.

Of the dynamics, the *narrative* element correlates with a strongly positive effect with the group of games we assigned to narratives, $p < .001$, $r_s = 0.93$. The Quiz games, on the other hand, correlate strongly negatively with the *narrative* dynamic, $p < .001$, $r_s = -0.76$.

Of the mechanics, narrative games and *resource acquisition* correlate with a large positive effect, $p = .001$, $r_s = 0.57$.

Of the group of components, a statistically significant correlation was demonstrated for two of the items with the narrative games: *avatar* correlated with a large positive effect, $p < .001$, $r_s = 0.93$, and *quest* also correlated with a large positive effect, $p < .001$, $r_s = 0.86$. The quiz games correlate with two of the components: *avatar* with a medium negative effect, $p < .001$, $r_s = -0.48$, and *quest* with a strong negative effect, $p = .001$, $r_s = -0.67$.

Following Schmidt-Kraepelin et al. [29], it is hypothesized that user ratings are influenced by the number of game design elements. To investigate whether there is a correlation between

the number of game design elements found and the user ratings, it was first checked whether the data are normally distributed. Both variables are not normally distributed according to the Shapiro-Wilk test: Sum of game design elements $p = .004$; user rating $p < .001$. Therefore, again a non-parametric test was used. The Spearman correlation does not become statistically significant, and we can assume that there is no correlation between the number of game design elements and user rating, $p = .103$, $r_s = -0.31$. Likewise, no statistically significant correlation could be found between the sum of game design elements and the number of user ratings, $p = .674$, $r_s = 0.08$.

5. Discussion

It is noticeable that the more specific the description of the individual elements of Werbach and Hunter [32] are, the more rarely they occur in our analysis. The dynamics and mechanics, which are described in general terms by Werbach and Hunter, occur more frequently in our analysis than the components. The components, which are specified very precisely, can be found only in a few, partly in none of the Alexa skills. An example of this is the *challenge* mechanic and the *boss fight* component. In the case of the *boss fight*, it is precisely defined as a specific *challenge* and in which form it occurs. Therefore, it is unsurprising that this precisely defined component only occurs in three games. While the mechanic *challenge* is far more generally defined and therefore was discovered much more frequently, specifically in 28 out of 29 games.

According to Werbach and Hunter [32], the higher levels of their model should be implemented by those below, as in the example above. However, this is not always the case in our sample. Thus, seven games appear in the sample that contain elements from the group of dynamics and mechanics but none of the components. This leads to the conclusion that either there are not enough components for audio games in Werbach and Hunter's model or that the three categories do not apply to audio games.

It is also intriguing that only one game included the dynamic *relationships*. This can be attributed to the fact that Alexa games are often single-player games. Therefore it is especially challenging to implement this dynamic. In our sample, the dynamic *relationships* could only be identified in the Alexa skill "Das Fußball

Tippspiel." In this game, the user can bet against another player in a real soccer match. The dynamic *relationships* is thus implemented through the mechanic *competition*. In the tested *narrative games*, the player's task consists of options to choose from or puzzles that lead the player through the narrative. There is no opponent, so the players only complete the tasks to go on. To implement *competition* in a voice-enabled game, for example, a fictional opponent could be introduced, or there could be the possibility to let several players interact with the game at once. This could be done, for example, by making turn-based decisions about how different players move through the story. Other elements that could be used to implement *relationships* could be considered, e.g., the components *combat*, *gift*, *leaderboard*, *team*, and *social graph*. These components appear sporadically in our sample. However, the implementation of the components was not very appealing, so no feeling of a relationship could be perceived during testing.

Several games implementing a competitive design were found in the sample, and on the other hand, not one cooperative game was identified. Cooperation seems to be a difficult concept to apply in the context of IVAs, which may explain why it was not found in our sample. Bräuer and Mazarakis [8] also attempted to compare a cooperative gamification approach in an IVA application with a competitive one. However, no differences could be found, which was attributed, among other things, to the fact that the cooperative and competitive aspects should be emphasized more clearly.

A remarkable aspect of our sample is the rare occurrence of game design elements like *badges* and *levels* [21], which are typically included in many games and gamified apps. *Badges* are digital artifacts that users receive for completing certain tasks [3]. They are explicitly represented visually [13]. This visual representation is not possible in voice-enabled games, which explains why the game design element does not appear in our sample. A non-visual alternative to *badges* is *achievements*, which are often equated with *badges* since both are awarded to the user for completing certain tasks [21]. However, in our sample, we do not find any games that include *achievements* either. It is possible that the implementation of *achievements* through spoken language is too distracting and is therefore not used in voice-enabled games.

Levels, which are also frequently used in other contexts, do not necessarily require a visual

component but are dependent on the player's progress [34]. *Levels* can be expressed in terms of missions that are completed to reach the end of the game, increasing difficulty, or an improvement in the player's experience and skills [19]. However, no form of progress was evident in most of the games studied. This lack of progress can be attributed to the brevity of most games.

Our study offers added value as a complement to the audio game framework of Ciccio and Quesada [15]. The framework provides a good basis to guide developers in creating audio games. However, the framework offers only a small set of game design elements. Building on our work, the list of five elements presented in the framework can be extended. For example, *quests*, *avatars*, *virtual goods*, or *points* could be dynamics that are well-suited to be used in voice-enabled games. Furthermore, it seems that all the mechanics can be implemented well, except for *cooperation*.

Unlike Schmidt-Kraepelin et al. [29], no correlation between the number of game design elements and the user rating of the Alexa skills could be demonstrated. This could be attributed to the fact that our sample consisted of games and not gamified mHealth apps. In a game, unlike a gamified application, the focus is not only on using game design elements but also on creating a gameful experience [16]. Thus, it can be assumed that a successful interaction of the elements has a stronger influence on how well a game is evaluated. Therefore, in gamified apps, it makes sense to consider whether the integration of more game design elements positively influences the evaluation of the application since the actual application should be complemented by adding game design elements. In gamification research, using multiple game design elements is not unusual but not entirely free of criticism [23].

A closer look at the two types of games identified during testing reveals how games for IVAs are currently implemented. By correlating the two types of games with the game design elements, implications can also be derived on how narrative and quiz games can be realized.

Narrative games are characterized by the *narrative* dynamic. There is a significant correlation with the game design element *avatar* as well. In *narrative games*, the player mostly moves through a game world. In this world, the player takes over a role that is embodied by an *avatar*. As this *avatar*, the player can act in the game world. The *avatar* usually has one or more *quests*, which the player should fulfill to reach the end of the game. This is also evident in the

positive correlation between *narrative games* and the *quest* element. A positive correlation among *narrative games* can also be seen in the game design element *resource acquisition*. The player sometimes acquires items in the game world that can be relevant later on in the narrative. An example is the skill "Escape Room," where the player can unlock a crowbar that can be used to gain access to further items in the game.

The negative link between *quiz games* and the *narrative* dynamic, as well as with the elements *avatar* and *quests*, supports the classification of games into these two types. In most *quiz games*, there is no game world for the players to explore. Consequently, designing an *avatar* to represent the player in the virtual world is not necessary. No *quests* are defined, only questions are asked and answered without a more profound mission. *Quiz games* usually seem to follow a constant pattern. First, a question is asked, and the player can answer. Afterward, the player gets *feedback* on whether the answer is right or wrong. If the answer is correct, the player sometimes gets a *reward*. It is also possible to punish the player for wrong answers, e.g., by deducting points. Then usually, a new *turn* begins with another question.

It is also noticeable that, according to our classification, significantly more games can be assigned to quizzes than specified in the Alexa Store categories (we assigned 13 compared to four in the Alexa store). In this sample, the division of games into the two types was very clear. Nevertheless, *quiz*, and *narrative games* need not always to represent two disjoint sets. For example, it would be perfectly possible to combine a quiz with a narrative.

The three games that cannot be assigned to one of the two types have some real-world reference. One is the previously described game "Das Fußball Tippspiel." The other two are old children's games: "Schnick Schnack Schnuck" (rock paper scissors), where the player can compete against Alexa or other users in the hand game by choosing orally one of the three shapes. In the game "Stopptanz," players dance around the room and must stop moving when the music stops. Whoever moves last is kicked out.

Most of the Alexa skills are relatively simple, which is not surprising given the technical limitations taken into account in IVA development. For example, push notifications cannot usually be sent to remind the player to come back. In addition to limiting the output to audio, designing the input is challenging, as all inputs are state-dependent. For example, inputting

"yes" in one state will cause the game to end, and in another state, the avatar will put on his armor.

6. Limitations and Future Research

A limiting factor of our study is the categorization, according to Werbach and Hunter [32]. The model was chosen due to its spread, relatively large number of elements, systematic structure, and the provision of descriptions of the elements. There is, however, a multiplicity of further models on the basis that one could try to classify voice-enabled games. Particularly from the category of components, some elements do not appear at all in our sample and may not fit the audio context, such as *badges*. Other models could possibly offer more components that are better implemented without visual support. For example, Thiebes et al. [31] mention audio feedback in addition to classical feedback. Koivisto and Hamari [20] identify other elements that can be implemented acoustically, such as timer, reminder, and virtual world.

Another limitation is the restriction to German-language Alexa skills. Voice assistants are more widely used in the USA than in Germany [11,12]. Accordingly, there are significantly more skills in the US Alexa Store. By looking at the English-language Alexa skills, a significantly larger sample of games could be examined.

In our study, the focus was put on Alexa skills that can be found among the top-rated ones to get an impression of which game design elements are particularly suitable. In further studies, games rated lower could also be considered for comparison. This could provide insights into which game design elements might be unsuitable and disrupt the game experience.

In video games auditory icons and earcons are additional options to convey information via audio [25]. These can also be used to illustrate information like the advancement in a level or the unlocking of a badge. A follow-up study could look at how these are used in IVA games.

Another way to expand on our study would be to examine the game types more. For example, it could be considered whether other game types can be identified in a larger sample, complementary to narrative or quiz games. In addition, the three games in our sample that have yet to be assigned could serve as a possible reference point.

One challenge our study faced was the sometimes very buggy implementation of the games, even though they were among the highest

ranked in the German Amazon Alexa store. Errors included outdated information in quiz games or technical problems that caused the games to crash. This made testing difficult since the games had to be started from the beginning repeatedly, which made playing the games challenging.

Another issue was that in some cases, Alexa did not understand every voice input correctly, or what was said could not be realized in the game. This resulted in some sections of games not being able to be explored or game variations only being minimally tested. Currently, many voice-enabled games have not yet met the expectations seen in science fiction examples [1]. However, the potential of voice-enabled games is already emerging, and through systematic investigation, the design can be steadily improved.

7. Conclusion

In our study, a systematic analysis of game design elements in voice-enabled games for Amazon Alexa was conducted. The analysis shows which game design elements are used how often. Based on this, initial implications can be derived as to which game design elements are particularly good to implement in the context of IVAs, such as *narrative*, *feedback*, *points*, or *quests*, and which ones rather not, such as *badges* and *levels*. The results give a first impression of how gamification could be designed in the context of IVAs. At the same time, the results show that IVA games are mostly designed as either quizzes or narratives. Further studies could investigate which other game design elements, especially from the field of audio games, are used in the context of IVAs.

8. References

- [1] Fraser Allison, Marcus Carter, Martin Gibbs, and Wally Smith. 2018. Design Patterns for Voice Interaction in Games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, ACM, Melbourne VIC Australia, 5–17. DOI:<https://doi.org/10.1145/3242671.3242712>
- [2] Fraser Allison, Joshua Newn, Wally Smith, Marcus Carter, and Martin Gibbs. 2019. Frame Analysis of Voice Interaction Gameplay. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, Glasgow Scotland Uk, 1–14.

- DOI:<https://doi.org/10.1145/3290605.3300623>
- [3] Judd Antin and Elizabeth F Churchill. 2011. Badges in Social Media: A Social Psychological Perspective. In *CHI 2011 Gamification Workshop Proceedings*, Vancouver, British Columbia, Canada, 1–4.
- [4] Allan de Barcelos Silva, Marcio Miguel Gomes, Cristiano André da Costa, Rodrigo da Rosa Righi, Jorge Luis Victoria Barbosa, Gustavo Pessin, Geert De Doncker, and Gustavo Federizzi. 2020. Intelligent Personal Assistants: A Systematic Literature Review. *Expert Systems with Applications* 147, (June 2020), 113193. DOI:<https://doi.org/10.1016/j.eswa.2020.113193>
- [5] Dennis Benner, Sofia Schöbel, Christoph Süess, Vera Baechle, and Andreas Janson. 2022. Level-Up your Learning – Introducing a Framework for Gamified Educational Conversational Agents. *Wirtschaftsinformatik 2022 Proceedings* (January 2022). Retrieved from <https://aisel.aisnet.org/wi2022/hci/hci/5>
- [6] Paul Best. 2022. Amazon Alexa on Track to Lose \$10 Billion This Year, Described as “Colossal Failure” in New Report. *FOXBusiness*. Retrieved November 24, 2022 from <https://www.foxbusiness.com/technology/amazon-alexa-track-lose-10-billion-year-described-colossal-failure-new-report>
- [7] Leonardo Bilic, Markus Ebner, and Martin Ebner. 2020. A Voice-Enabled Game Based Learning Application using Amazon’s Echo with Alexa Voice Service: A Game Regarding Geographic Facts About Austria and Europe. *Int. J. Interact. Mob. Technol.* 14, 03 (February 2020), 226. DOI:<https://doi.org/10.3991/ijim.v14i03.12311>
- [8] Paula Bräuer and Athanasios Mazarakis. 2022. “Alexa, Can We Design Gamification Without a Screen?” - Implementing Cooperative and Competitive Audio-Gamification for Intelligent Virtual Assistants. *Computers in Human Behavior* 135, (October 2022), 107362. DOI:<https://doi.org/10.1016/j.chb.2022.107362>
- [9] Paula Bräuer and Athanasios Mazarakis. 2022. How to Design Audio-Gamification for Language Learning with Amazon Alexa?—A Long-Term Field Experiment. *International Journal of Human-Computer Interaction* (December 2022), 1–18. DOI:<https://doi.org/10.1080/10447318.2022.2160228>
- [10] Bret Kinsella. 2020. Global Smart Speaker Growth Cools in Q1 as Pandemic Leads to Declining China Sales, Amazon Retains Top Spot Says Strategy Analytics. *Voicebot.ai*. Retrieved November 28, 2022 from <https://voicebot.ai/2020/05/25/global-smart-speaker-growth-cools-in-q1-as-pandemic-leads-to-declining-china-sales-amazon-retains-top-spot-says-strategy-analytics/>
- [11] Bret Kinsella. 2021. U.S. Smart Speaker Growth Flat Lined in 2020. *Voicebot.ai*. Retrieved December 4, 2022 from <https://voicebot.ai/2021/04/14/u-s-smart-speaker-growth-flat-lined-in-2020/>
- [12] Bret Kinsella. 2021. Germany Smart Speaker Adoption Closely Mirrors U.S. Pattern - New Report with 30+ Charts. *Voicebot.ai*. Retrieved December 4, 2022 from <https://voicebot.ai/2021/06/17/germany-smart-speaker-adoption-closely-mirrors-u-s-pattern-new-report-with-30-charts/>
- [13] Sven Charleer, Jose Luis Santos, Joris Klerkx, and Erik Duval. 2014. Improving Teacher Awareness Through Activity, Badge and Content Visualizations. In *New Horizons in Web Based Learning*, Yiwei Cao, Terje Väljataga, Jeff K.T. Tang, Howard Leung and Mart Laanpere (eds.). Springer International Publishing, Cham, 143–152. DOI:https://doi.org/10.1007/978-3-319-13296-9_16
- [14] Vanessa Wan Sze Cheng, Tracey Davenport, Daniel Johnson, Kellie Vella, and Ian B Hickie. 2019. Gamification in Apps and Technologies for Improving Mental Health and Well-Being: Systematic Review. *JMIR Ment Health* 6, 6 (June 2019), e13717. DOI:<https://doi.org/10.2196/13717>
- [15] José Antonio Cicció and Luis Quesada. 2018. Framework for Creating Audio Games for Intelligent Personal Assistants. In *Advances in Human Factors in Wearable Technologies and Game Design* (Advances in Intelligent Systems and Computing), Springer International Publishing, Cham, 204–214. DOI:https://doi.org/10.1007/978-3-319-60639-2_21
- [16] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining “Gamification.” In *Proceedings of*

- the 15th international academic MindTrek conference: Envisioning future media environments*, ACM Press, 9–15. DOI:<https://doi.org/10.1145/2181037.2181040>
- [17]Ulrike Hammerschall. 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)*, IEEE, Dubai, United Arab Emirates, 95–101. DOI:<https://doi.org/10.1109/EDUCON.2019.8725251>
- [18]Matthew B. Hoy. 2018. Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly* 37, 1 (January 2018), 81–88. DOI:<https://doi.org/10.1080/02763869.2018.1404391>
- [19]Karl M. Kapp. 2012. *The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education*. John Wiley & Sons, New York, NY, USA.
- [20]Jonna Koivisto and Juho Hamari. 2019. The Rise of Motivational Information Systems: A Review of Gamification Research. *International Journal of Information Management* 45, (2019), 191–210. DOI:<https://doi.org/10.1016/j.ijinfomgt.2018.10.013>
- [21]Jeanine Krath, Linda Schürmann, and Harald F. O. von Korflesch. 2021. Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in Human Behavior* 125, (December 2021), 106963. DOI:<https://doi.org/10.1016/j.chb.2021.106963>
- [22]J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174. DOI:<https://doi.org/10.2307/2529310>
- [23]Athanasios Mazarakis. 2021. Gamification Reloaded: Current and Future Trends in Gamification Science. *i-com* 20, 3 (November 2021), 279–294. DOI:<https://doi.org/10.1515/icom-2021-0025>
- [24]Michael McTear, Zoraida Callejas, and David Griol. 2016. *The Conversational Interface*. Springer International Publishing, Cham. DOI:<https://doi.org/10.1007/978-3-319-32967-3>
- [25]Patrick Ng and Keith Nesbitt. 2013. Informative sound design in video games. In *Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death*, ACM, Melbourne Australia, 1–9. DOI:<https://doi.org/10.1145/2513002.2513015>
- [26]Cathy Pearl. 2016. *Designing Voice User Interfaces: Principles of Conversational Experiences*. O’Reilly Media, Inc.
- [27]Mikko Rajanen and Dorina Rajanen. 2017. Usability Benefits in Gamification. In *Proceedings of the 1st International GamiFIN Conference*, Pori, Finland, 87–95.
- [28]Niklas Röber and Maic Masuch. 2005. Leaving the Screen New Perspectives in Audio-Only Gaming. (July 2005). Retrieved March 18, 2021 from <https://smartech.gatech.edu/handle/1853/50168>
- [29]Manuel Schmidt-Kraepelin, Scott Thiebes, and Ali Sunyaev. 2019. Investigating the Relationship Between User Ratings and Gamification – A Review of mHealth Apps in the Apple App Store and Google Play Store. *Hawaii International Conference on System Sciences 2019 (HICSS-52)* (January 2019). Retrieved from <https://aisel.aisnet.org/hicss-52/da/gamification/7>
- [30]Manuel Schmidt-Kraepelin, Scott Thiebes, Minh Chau Tran, and Ali Sunyaev. 2018. What’s in the Game? Developing a Taxonomy of Gamification Concepts for Health Apps. *Proceedings of the 51st Hawaii International Conference on System Sciences* (2018), 1217–1226.
- [31]Scott Thiebes, Sebastian Lins, and Dirk Basten. 2014. Gamifying Information Systems - a Synthesis of Gamification Mechanics and Dynamics. *ECIS 2014 Proceedings* (June 2014), 1–17.
- [32]Kevin Werbach and Dan Hunter. 2012. *For the Win: How Game Thinking Can Revolutionize Your Business*. Wharton Digital Press, Philadelphia, PA, USA.
- [33]Melvyn W.B. Zhang, Jiang Bo Ying, Guo Song, and Roger C.M. Ho. 2018. A Review of Gamification Approaches in Commercial Cognitive Bias Modification Gaming Applications. *THC* 26, 6 (December 2018), 933–944. DOI:<https://doi.org/10.3233/THC-181313>

[34]Gabe Zichermann and Christopher Cunningham. 2011. *Gamification by Design: Implementing Game Mechanics in Web and*

Mobile Apps. O'Reilly Media, Sebastopol, CA, USA.

9. Appendix

Table 2: List of Alexa skills with ratings and categorization according to the Amazon Alexa Store and our categorization.

| Name of the skill | Rating | | Game type by Amazon | | Game type | |
|--|--------|-------|---------------------|------------------|-----------|------|
| | Stars | Count | Games | Knowledge & Quiz | Narrative | Quiz |
| Das Fußball Tippspiel | 5 | 23 | 1 | 0 | 0 | 0 |
| Escape Room | 4.5 | 599 | 1 | 0 | 1 | 0 |
| Twist Tale | 4.5 | 23 | 1 | 0 | 1 | 0 |
| TKKG-Mein Abenteuer | 4 | 27 | 1 | 0 | 1 | 0 |
| Love Stories | 4.5 | 20 | 1 | 0 | 1 | 0 |
| Quiz des Tages | 4 | 10626 | 0 | 1 | 0 | 1 |
| Kinderquiz Richtig oder Falsch? | 4.5 | 2698 | 1 | 0 | 0 | 1 |
| Star Commander | 4 | 52 | 1 | 0 | 1 | 0 |
| Japan Quiz | 5 | 7 | 1 | 0 | 0 | 1 |
| Camp Fires | 5 | 6 | 1 | 0 | 1 | 0 |
| Meine Ferien auf dem Martinshof | 4 | 1563 | 1 | 0 | 1 | 0 |
| Schwäbisch Quiz mit Schwäbman von Antenne 1 | 4.5 | 44 | 0 | 1 | 0 | 1 |
| Rittermanager | 4.5 | 550 | 1 | 0 | 1 | 0 |
| Ja, Hoheit | 4 | 181 | 1 | 0 | 1 | 0 |
| Falsch ist Richtig- Das total Verrückte Quiz | 4.5 | 13 | 1 | 0 | 0 | 1 |
| Das Rätsel des Tages | 4 | 7135 | 1 | 0 | 0 | 1 |
| Limostand | 4 | 2397 | 1 | 0 | 1 | 0 |
| Heldentum für Anfänger | 4.5 | 14 | 1 | 0 | 1 | 0 |
| Twenty Questions | 3.5 | 19 | 1 | 0 | 0 | 1 |
| Nationalhymnen raten | 5 | 4 | 0 | 1 | 0 | 1 |
| Wörterfuchs | 5 | 4 | 1 | 0 | 0 | 1 |
| Schnick Schnack Schnuk | 4 | 4991 | 1 | 0 | 0 | 1 |
| Stopptanz | 4 | 2888 | 1 | 0 | 0 | 0 |
| Mein Königreich | 4 | 1525 | 1 | 0 | 0 | 0 |
| Der Zauberwald | 3.5 | 235 | 1 | 0 | 1 | 0 |
| Tierspiel | 4 | 992 | 1 | 0 | 1 | 0 |
| Knoppers Quiz | 4 | 463 | 1 | 0 | 0 | 1 |
| Das Ultimative Hip Hop Quiz | 5 | 9 | 0 | 1 | 0 | 1 |
| Akinator | 4 | 16767 | 1 | 0 | 0 | 1 |

CHAPTER

SEVEN

General Discussion

Gamification is a concept used in many different contexts to motivate users. This thesis has shown that investigating whether gamification can be implemented for IVAs is an innovative and complex question. In this chapter, the results of the four studies of this thesis are related to each other in order to address the main research gap (Does audio-gamification impact motivation and performance in the context of Intelligent Virtual Assistants?). The implementation possibilities for audio-gamification are manifold (Bräuer et al., 2023), although the interaction possibilities with an IVA seem to be limited due to the lack of a screen. The four studies provided insight into the opportunities for gamification even without visual game design elements. Theoretical and practical implications will be derived from the results, and an outlook will be given on future research approaches based on this thesis.

7.1. Theoretical Implications

Due to the novelty of using gamification without visual elements, the research question addressed by this thesis is already pioneering. In addition, this thesis opens up a wide field of possibilities for the combination of the two research areas of gamification and IVAs. The conceptual work in this thesis shows that gamification is not a concept that necessarily requires a visual component. While the definition of gamification does not limit the choice of game design elements (Deterding et al., 2011; Huotari & Hamari, 2017; Zichermann & Cunningham, 2011), only visual elements are applied in practice and most studies. This work rethinks the concept of gamification and applies it to the previously

neglected audio level. A definition of audio-gamification complements this conceptual understanding and provides a basis for empirical arguments.

Contribution to Theoretical Understanding of Gamification

From the perspective of gamification research, this thesis contributes to new theoretical insights, especially with regard to gamification design. Concerning RQ1 (How do different game design elements affect motivation?), the effect of individual game design elements is systematically investigated. This extends existing research in the field of gamification (Mekler et al., 2017). In both studies (chapter 3 and chapter 5), different game design elements' effects were analyzed to draw conclusions about their individual effects. Both studies indicate that gamification does not require complex constructs consisting of a multitude of game design elements. Individual elements can be sufficient to motivate users. However, not all game design elements have the same effect. The two studies demonstrate that the impact of game design elements depends on their implementation and the context in which they are used. These findings provide a crucial foundation for gamification research, which presently demands customization in gamification concepts (Rodrigues et al., 2020). Looking at gamification at the level of individual game design elements is essential for customization, as they are adapted according to user profiles based on factors such as age, gender, or player type (Klock et al., 2020).

An open question in the gamification literature is how persistent the effects of gamification are on user motivation and performance (Nacke & Deterding, 2017; Thomas et al., 2022). Do only short interventions work, as most laboratory studies have investigated, or can a motivational effect be sustained over time? Study 3 is one of the few studies to examine gamification in practice and over time. Data were collected for one year. The experimental setting allowed users to interact with Alexa skills at their discretion. Statistical analysis shows that audio-gamification increased user motivation within a single session but did not affect long-term skill use. The implementation of audio-gamification did not encourage users to start the skill more often. This leads to the assumption that most users are motivated by (audio-) gamification in the short term but that the effect does not continue after the application has been used. However, it should be noted that retention of IVA skills is very low (Planner, 2018). This means that a possible effect of audio-gamification on long-term use may not be detectable because

other factors prevent the user from using it again. What these factors might be remains an open question.

An essential insight for gamification research from this thesis is that gamification is not a concept that necessarily has to use visual elements. Previous research in various contexts has shown that gamification is an effective means to increase user motivation. However, the game design elements used in this context have so far been implemented visually, as shown by various reviews summarizing which game design elements are used in different studies (Klock et al., 2020; Koivisto & Hamari, 2019; Saleem et al., 2022). The approach of using audio-gamification to increase motivation has proven successful in the studies presented in this thesis. This allows a new audio layer to be added to the theoretical concept of gamification. This new research layer provides game design elements whose effects on theoretical concepts such as SDT, flow, or goal setting have not been considered. These audio elements can be considered individually, as was done here in the context of IVAs, or in combination with visual game design elements.

Contribution to Theoretical Understanding of IVAs

This thesis's empirical work focuses on transferring the gamification concept into the context of IVAs. Previously, it was assumed that interactions with IVAs should be designed to be as streamlined as possible to not overwhelm the user (Murad et al., 2018). However, design patterns and heuristics for VUI are much less developed than GUI design resources (Murad et al., 2021). In theory, a streamlined IVA design would be countered by the integration of gamification. However, studies 2 and 3 show that integrating game design elements has positive effects. It can be assumed that a voice-based interface can also benefit from more complex design concepts, such as audio-gamification.

One challenge of IVA design that can benefit from more complex design approaches is the implementation of feedback (Murad et al., 2021; Myers et al., 2018). A fundamental principle of gamification is the implementation of feedback (Zichermann & Cunningham, 2011), e.g., through progress indicators such as levels or progress bars. The analysis of the game design elements and the IVA games from Study 4 provides an overview of which known feedback mechanisms of gamification can also be applied in the audio context. However, many common elements are not found in the analyzed IVA games.

The results of Study 3 also show that not every form of feedback is effective without a screen.

The range of games available for IVAs is still very limited but has increased significantly in recent years with the proliferation of smart speakers (Bräuer et al., 2023). The IVA games analyzed in Study 4 mostly correspond to one of two schemes: quiz or narrative-based games. Both types of games can be implemented well through speech and are similar to well-known concepts such as radio quiz shows or text adventures. In particular, the large proportion of quiz games is consistent with the results of other analyses (Allison et al., 2018). By looking at the games in detail on the level of game design elements, it becomes clear that not only the spectrum of games is limited, but also the choice of game design elements used. It seems that IVA games, and thus the game design elements that can be used in the context of IVAs, represent a small subset of the diverse set of elements used in video games, e.g., challenges or narrative.

In addition to the theoretical insights that can be derived from the results of the studies, there is also an implication for the further development of IVAs in general. The rapid development of language models such as ChatGPT suggests that in the near future, IVAs will be much more flexible in responding to user requests (Lee, 2023). Will IVAs be more human-like with whom “normal conversations” just like with other humans will be possible? Or will it move away from this approach to try to design machines like humans and continue to communicate with machines in our own way? Uncanny valley theory describes this conflict between humanity and machine-like appearance (Mori et al., 2012). It assumes that the reaction on a human-like robot at a certain point suddenly turns negative if it approaches an alife-like appearance but does not reach it. Garcia and Lopez (2020) showed that this concept can also be found in the context of IVAs. How the user perceives an IVA depends on several different factors (Ling et al., 2021). When exploring an IVA, for example, playful approaches are often found, such as asking for a joke or singing a song, which has an influence on the personification of the IVA (Luger & Sellen, 2016). Depending on the goal of an IVA (a human-like assistant or a machine-like one), the integration of game design elements can help to support or counteract this goal. Game design elements such as narration can also enhance the impression of having a conversation with a more human IVA, e.g. by shaping the narrative through a narrator. Classic game design elements such as points, badges, and leaderboards are more likely to counteract this and lead to a more machine-like representation. Even more than is already

necessary for the design of gamified applications, the focus in developing gamified IVAs will have to be on a suitable implementation and a careful selection of game design elements. The thesis of audio-gamification creates new possibilities to design IVA applications in both directions of the spectrum. Through a thoughtful design of a gamification concept, the playful aspects should create either more human interactions or more machine-like dialogs.

7.2. Practical Implications

The conceptual contribution of this thesis shows that, complementary to previous understanding and practice, gamification is a concept that can be implemented through audio elements. From the empirical results, some practical implications can be derived, which can be helpful for the development of gamified applications as well as for the design of IVA applications.

Contribution to (Audio-) Gamification Design

Adding gamification to an application is not an easy task. To develop a suitable concept, the designer needs experience and has to get to know the application to be gamified, which is time- and cost-intensive. Study 1 showed that the integration of individual game design elements can be sufficient to increase user motivation. Suppose the gamification concept is designed appropriately for the context. In that case, it is not necessary to create an infinitely complex construct consisting of different game design elements, but one can focus on individual elements. This insight also clarifies the difference between gamification and (serious) games. Gamification is not about turning an application into a game. For the implementation of gamification in practice, the recommendation can be derived from this work to select individual game design elements that fit into the context of the application and are tailored to the target group. In this way, more straightforward concepts can save resources during development and be used effectively.

The frameworks available in gamification research, such as the one used in Study 4 by Werbach and Hunter (2012), offer various game design elements. Adapting these elements to different contexts and personality types is a challenging task. The studies conducted in this thesis show that gamification cannot be implemented 1:1 in audio contexts. In addition, study 3 showed that users were not motivated by the integration of speech-implemented achievements. One possible reason for this is that in visually implemented scenarios, this game design element can be displayed parallel to the actual

application, e.g., as a list in the corner of the screen. In a purely sound-based implementation, the achievements become too prominent and distract the user from the actual activity. The analysis of the Alexa games in Study 4 supports the assumption that typical gamification game design elements, such as badges and achievements, are not easily transferred to audio-gamification. Badges or achievements were not found in any of the Alexa games studied. Developers of Alexa games seem to rely on elements other than those that can be easily integrated into GUIs. In practice, whether existing approaches can be implemented without a screen must be carefully considered. To facilitate the implementation of audio-gamification, it is therefore recommended to look at existing gamification frameworks as well as examples of gamified interactions without a screen, such as those presented in chapter 2.3.

IVAs also have great potential for individuals with visual impairments (Abdolrahmani et al., 2018). The now relatively wide distribution of IVAs contributes to the platforms' continuing to develop rapidly. In order not to diminish the potential of IVAs for these user groups, further development should seek to enhance the strength of this technology, i.e., interaction without a screen. Audio-gamification offers new possibilities for developing applications for IVAs without limiting the advantages of this technology. By developing audio-gamification, a wide variety of IVA applications can be made more motivating, such as learning tools or smart home applications for energy conservation. Gamified applications can be accessed by individuals with visual impairments by avoiding the use of a screen.

Contribution to IVA Design

Many skills for IVAs are currently basic and not very motivating. Even if a skill is designed to be used repeatedly and retained over a long period of time, the very low retention rate of about 6% shows that this is not easy to achieve (Planner, 2018). The concept of audio-gamification presented in this thesis could address this weakness. How this could be implemented in practice is shown by two examples.

An important area for smart speaker manufacturers is online shopping via IVA (Rzepka et al., 2020), also referred to as voice shopping or voice commerce (Mari, 2019). IVAs offer the ability to provide users with product recommendations at home (Rhee & Choi, 2020), information about new products, and direct purchases via voice command. Here, too, the shopping experience could be enhanced through audio-gamification. For

example, instead of simply reading product descriptions, the IVA could provide a kind of interactive "show" that presents product features, benefits, and even reviews entertainingly. Users could ask questions to learn more about a product. To improve product recommendations, the IVA could take a fun quiz with the user to better understand preferences, needs, and budgets. The assistant could make customized product suggestions based on the user's answers.

Another related application area where the results of this work could potentially be helpful is in virtual and augmented reality applications. Many of the headsets currently in use also offer the ability to provide voice input, such as the HoloLens 2, PICO, or Meta Quest. Since many VR and AR use cases benefit from the user having their hands free, such as during surgery with an AR headset, interacting with the application via voice is a practical option (Monteiro et al., 2021). Combining gamification with AR applications has already proven valid (Bräuer & Mazarakis, 2019). However, the impact of audio in AR and VR has been poorly explored (Bosman et al., 2023). Providing game design elements purely acoustically could increase the potential of VR and AR applications without overloading the environment with additional visual stimuli.

Ethical Impact of Audio-Gamification

Ethical concerns should also be considered when researching and implementing audio-gamification. First, it should be noted that IVAs are often criticized today because they can potentially collect information unintentionally from the user (Lau et al., 2018). Since IVAs have to actively analyze sounds at all times to be able to react to a user's wake word at any time, users could be monitored almost continuously. Second, gamification is under criticism because it offers potential for abuse in addition to its positive effects. In addition to problems such as creating pressure in the work environment through comparison with others (Shahri et al., 2014), gamification can also be used to entice users to do things they would not do under normal circumstances (Thorpe & Roper, 2019). For example, gamification could trick users into giving personal information to an IVA they would generally have preferred to keep private. Another example is AI characters, who motivate players to make in-game purchases to improve their chances of winning (Austin, 2023).

Addressing such ethical concerns requires careful consideration and awareness. A few common approaches can help address these. One important approach is to collect and store user data securely. Attempts should be made to limit data collection to the minimum

necessary for functionality. This approach is called “privacy by design.” What data is collected should be transparently communicated. Users should be able to control the collection of their data and make informed decisions about its use. This helps to protect users from unwanted data collection, but can also be used to mark in-app purchases.

When designing gamification, developers should be aware of the psychological effects that can arise, such as the build-up of social pressure (Al-Msallam et al., 2023). To counteract this, care should be taken to balance the metrics used, both in the initial phase and in productive use. Feedback options for evaluating the application can also provide helpful information.

By considering these solutions, researchers and developers can contribute to a more ethical and responsible approach to audio gaming by minimizing potential risks and protecting the privacy and well-being of users.

7.3. Future Research Directions

This thesis aimed to determine the potential of audio-gamification to motivate users of IVAs. The research yielded an answer to the overarching question, highlighting new prospects for future investigations on various levels. The subsequent section indicates a few conceivable pathways.

The current hype around the ChatGPT language model brings criticism of the current state of development of Siri, Amazon Alexa, and Google Assistant but simultaneously opens up new opportunities (Lee, 2023). Language models such as ChatGPT and IVAs are based on two types of AI. ChatGPT is based on a so-called large language model, a system that can recognize and generate text on a massive dataset from the web. IVAs, on the other hand, use command-and-control systems that can understand and respond to a list of different questions and commands. If a user’s input is not in the list, the assistant just says it can’t help (Chen et al., 2023). Nevertheless, IVAs offer several advantages through spoken language interaction (Bouزيد & Ma, 2022), making it likely that the concept of voice control will become more prevalent in the coming years. The two AI-supported concepts, i.e., IVAs and language models like ChatGPT, are expected to be combined soon (Chen et al., 2023). Combining these two models has significant promise if the informative advantages of language models and the interactional advantages of IVAs are exploited wisely. The four studies outlined in this thesis illustrate the potential for audio-gamification to enhance motivation toward IVA usage. More customized and

adaptable interactions become feasible when coupled with speech models like ChatGPT. There is significant research potential in this area, which could be explored by various disciplines such as linguistics or UX design.

A starting point to directly follow up on this work would be to extend the analysis of game design elements from those typical of gamification to specific audio elements. Both the experiments with Alexa and the analysis of the Alexa games applied and examined game design elements found in well-known gamification frameworks. The mixed results of the experiments and the analysis of the Alexa games indicate that these frameworks are not directly applicable to audio-gamification. One way to approach this from a different research direction would be to take a look at audio games. The fourth study was a first step in this direction by analyzing Alexa games. However, the field of audio games could be expanded to include PC and smartphone games. Also, non-verbal audio elements such as music, auditory icons, or earcons have not been studied. An auditory icon is a characteristic sound associated with a real event (Jørgensen, 2006), such as clapping hands or dripping water. Earcons, conversely, are symbolic, more abstract sounds that cannot be automatically classified, such as a jingle played to signal that a new level has been reached (Jørgensen, 2006). These elements could be an important extension to existing gamification frameworks for the implementation of audio-gamification. Combining the existing frameworks with these specific audio elements could create a separate framework for audio-gamification. This would be an essential extension of existing design recommendations, especially for practical implementation. An investigation in combination with visual elements would also be interesting to explore audio elements. While audio in video games is often thought of as a means to convey mood, it has been shown that the usability of games is also influenced by the audio component (Jørgensen, 2008). The audio channel can convey additional information, such as an enemy player sneaking up on you. As mentioned above, audio has received little attention in gamification research. To gain a better understanding of the impact of audio-gamification and its effect on motivation, it is important to consider not only the effect of visual and audio elements but also their combination. Do audio-visual elements work better than elements that address only one level?

The nature of human-machine interaction via speech distinguishes audio-gamification development for IVAs from other contexts. Unlike using a PC or smartphone, multiple users can interact with an IVA at once (Winkler et al., 2019). Because a system can be

trained on different voices, individual users can be distinguished when using an IVA simultaneously (*What Is Alexa Voice ID?*, o. J.). Gamification can be implemented in various ways, but most often, solutions are developed and explored that a user can view individually through a personal interface. Sharing an endpoint device such as an IVA could also rethink design approaches to gamification. Future research could explore how sharing a gamified IVA or other device differs from private use. Collaborative and complementary gamification approaches could also support this distinction between individual and simultaneous use.

Not only may the interaction between different users be influential, but the relationship with an IVA (e.g., IVAs as companions) is also considered relevant to its use (Ling et al., 2021). One way to look at this relationship between humans and IVAs would be to examine the fulfillment of the need for social relatedness, according to SDT. Gamification frameworks offer various game design elements that can address the need for social relatedness, for example, through narratives (Sailer et al., 2017). In particular, collaborative approaches seem promising to meet this need (Dindar et al., 2021; Xi & Hamari, 2019), where users are encouraged to work together in teams or exchange and share. In Study 2, positive results were obtained with both cooperative and competitive audio-gamification. In particular, competitive elements seem to translate well to IVAs, as demonstrated by the analysis in Study 4. Future research could examine in more detail how game design elements affect the user's relationship with the IVA and how meeting basic psychological needs can influence this relationship.

From a methodological standpoint, this thesis holds potential for further exploration. All empirical studies of this thesis work with quantitative data, which were systematically collected in different settings. Using a between-subjects design allows for capturing user behavior without influencing the user, as can happen with self-reported information. The results show differences in user motivation when (audio-) gamification is integrated. Building on these results, a follow-up study could survey users to investigate whether and how they consciously perceive the integrated game design elements. Since there are hardly any recommendations for IVA game developers and no recommendations for audio-gamification, a qualitative approach could provide insights that contribute to developing design guidelines for audio-gamification.

CHAPTER

EIGHT

Conclusion

Technological advances are making human-machine interaction via speech possible. However, the ability to process speech alone does not seem to be enough to convince users of this promising technology. IVAs are currently criticized for being “dumb as a rock” (Chen et al., 2023), as the interaction usually does not live up to expectations. In the paradigm of making applications more interesting for IVAs, a promising research gap has been identified that will be filled through the use of gamification. This thesis makes a pioneering contribution to gamification and IVA research by considering how gamification can motivate users of IVAs. The empirical work in this thesis is the first to address this topic. Based on the insight that gamification can be implemented with a single game design element, several experiments were conducted to show how audio-gamification increases IVA users' motivation. In addition to explicitly answering the research questions of this thesis and thus building knowledge, the original empirical work of this thesis has drawn attention to the complexity of audio-gamification and the need for solid, scientifically based knowledge on this topic.

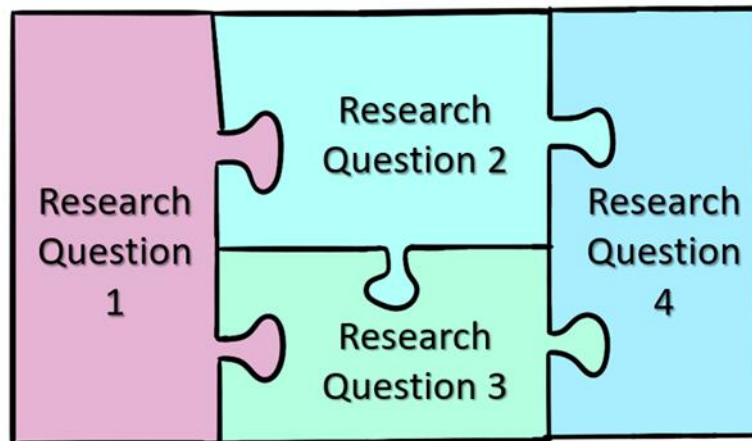


Figure 9 The pieces of the puzzle that fit together to illustrate the connection between the four research questions.

The results of this thesis provide a first insight into this wildly complex topic, in which a lot of follow-up research can still be done. With regard to the research questions posed at the beginning, the picture presented now fits together. The first research question was to examine individual game design elements to determine their specific influence on motivation. The first study's results show that each game design element examined - progress bar, narrative, feedback, and badges - has significant motivational effects. This finding is important because it shows that even integrating a single game design element can be sufficient to increase user motivation. This set the stage for research questions 2 and 3.

Research question 2, *“If and how does audio-gamification affect user motivation and performance?”* was addressed in the second study. It examined the use of audio-gamification in conjunction with IVAs to complete household tasks in a laboratory experiment. The integration of game design elements, such as points and levels via audio, led to a significant improvement in participants' attitudes and task completion time. This demonstrates the enormous potential of audio-gamification in the field of voice assistants and shows that visual representations of progress and feedback are not essential to achieving a motivational effect.

The third study combined gamification with IVAs for foreign language learning in a field experiment to answer research question 3, *“How do different game design elements in the context of IVAs affect motivation and performance of real-world subject behavior?”* It was found that audio-gamification positively impacted the number of translated vocabulary words and learning success, especially in the form of a leaderboard. This demonstrates the potential of audio-gamification to motivate learners to actively interact

with IVAs in a relevant educational context. These findings can provide essential impetus for the design of IVA-based educational applications.

The investigation of research questions 2 and 3 indicates that audio-gamification can motivate users to use an IVA. However, not every gamification approach that is successful in other contexts will have the same effect in the context of IVAs.

To complete the picture, research question 4, “*What are the dominant game design elements used in voice-enabled games for Amazon Alexa?*” was addressed in the final study. The fourth study examined the implementation of game design elements in voice-enabled games on the Amazon Alexa platform. The analysis revealed two identifiable game categories, quiz and narrative games, correlated with specific game design elements. The results suggest that traditional gamification frameworks do not translate one-to-one to audio-gamification. This indicates that particular factors must be considered when designing voice-enabled games to fully realize this platform's potential.

In summary, the four studies show that both gamification and audio-gamification can effectively increase motivation and engagement in different application domains. Considering game design elements individually makes selecting and using effective elements in a specific setting easier. Integrating gamification into IVAs opens up new possibilities for educational and everyday applications. However, the studies also illustrate that the design of audio-gamification poses specific challenges and that a differentiated consideration of the target group and the application context is essential to achieve optimal effects. Further research is warranted to better understand the mechanisms and relationships between game design elements, audio-gamification, and motivation and to further optimize the design of gamified IVAs.

References

- Abdolrahmani, A., Kuber, R., & Branham, S. M. (2018). „Siri Talks at You“: An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, 249–258. <https://doi.org/10.1145/3234695.3236344>
- Allison, F., Carter, M., Gibbs, M., & Smith, W. (2018). Design Patterns for Voice Interaction in Games. *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 5–17. <https://doi.org/10.1145/3242671.3242712>
- Al-Msallam, S., Xi, N., & Hamari, J. (2023). Ethical Considerations in Gamified Interactive Marketing Praxis. In C. L. Wang (Hrsg.), *The Palgrave Handbook of Interactive Marketing* (S. 963–985). Springer International Publishing. https://doi.org/10.1007/978-3-031-14961-0_41
- Antin, J., & Churchill, E. F. (2011). Badges in Social Media: A Social Psychological Perspective. *CHI 2011 Gamification Workshop Proceedings*, 1–4.
- Austin, M. L. (2023, März 24). „Toying with Alexa: Playing Music Games with Virtual Assistants“: Ludo2023: Twelfth Annual Conference on Video Game Music and Sound. *Ludo2023: Twelfth Annual Conference on Video Game Music and Sound*.
- Bartle, R. (1996). *Hearts, Clubs, Diamonds, Spades: Players Who Suit Muds*.
- Bicen, H., & Kocakoyun, S. (2018). Perceptions of Students for Gamification Approach: Kahoot as a Case Study. *International Journal of Emerging Technologies in Learning*, 13(2). <https://doi.org/10.3991/ijet.v13i02.7467>
- Bosman, I. D. V., Buruk, O. ‘Oz’, Jørgensen, K., & Hamari, J. (2023). The effect of audio on the experience in virtual reality: A scoping review. *Behaviour & Information Technology*, 1–35. <https://doi.org/10.1080/0144929X.2022.2158371>
- Bouزيد, A., & Ma, W. (2022). *The Elements of Voice First Style*. O’Reilly Media, Inc.

- Bräuer, P., & Mazarakis, A. (2019). Badges or a Leaderboard? How to Gamify an Augmented Reality Warehouse Setting. In J. Koivisto & Hamari (Hrsg.), *Proceedings of the 3rd International GamiFIN Conference—GamiFIN 2019* (Bd. 2359, S. 229–240). CEUR-WS.org.
- Bräuer, P., & Mazarakis, A. (2022a). “Alexa, Can We Design Gamification Without a Screen?”—Implementing Cooperative and Competitive Audio-Gamification for Intelligent Virtual Assistants. *Computers in Human Behavior*, *135*, 107362. <https://doi.org/10.1016/j.chb.2022.107362>
- Bräuer, P., & Mazarakis, A. (2022b). How to Design Audio-Gamification for Language Learning with Amazon Alexa?—A Long-Term Field Experiment. *International Journal of Human–Computer Interaction*, *0*(0), 1–18. <https://doi.org/10.1080/10447318.2022.2160228>
- Bräuer, P., Werner, F., & Mazarakis, A. (2023). Extracting Game Design Elements from Voice-Enabled Games: A Review of Amazon Alexa Skills. *Proceedings of the 7th International GamiFIN Conference*, 47–56.
- Caillois, R. (2001). *Man, Play, and Games*. University of Illinois Press.
- Cermak-Sassenrath, D. (2019). Current Challenges in Gamification Identified in Empirical Studies. In R. Ørngreen, M. Buhl, & B. Meyer (Hrsg.), *Proceedings of the 18th European Conference on e-Learning (ECEL)* (S. 119–127). Academic Conferences and Publishing International Limited.
- Chen, B. X., Grant, N., & Weise, K. (2023, März 15). How Siri, Alexa and Google Assistant Lost the A.I. Race. *The New York Times*. <https://www.nytimes.com/2023/03/15/technology/siri-alex-google-assistant-artificial-intelligence.html>
- Cho, M., Lee, S., & Lee, K.-P. (2019). Once a Kind Friend is Now a Thing: Understanding How Conversational Agents at Home are Forgotten. *Proceedings*

- of the 2019 on Designing Interactive Systems Conference, 1557–1569.
<https://doi.org/10.1145/3322276.3322332>
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition* (S. 127–149). American Psychological Association.
<https://doi.org/10.1037/10096-006>
- Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., Earley, D., & Bandeira, N. (2017). „What Can I Help You with?“. Infrequent Users' Experiences of Intelligent Personal Assistants. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–12. <https://doi.org/10.1145/3098279.3098539>
- Csikszentmihalyi, M. (2009). *Flow: The Psychology of Optimal Experience*. Harper Collins.
- Csikszentmihalyi, M. (2014). *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*. Springer Netherlands.
<https://doi.org/10.1007/978-94-017-9088-8>
- Da Silva, T. H. O., Furtado, V., Furtado, E., Mendes, M., Almeida, V., & Sales, L. (2022). How Do Illiterate People Interact with an Intelligent Voice Assistant? *International Journal of Human-Computer Interaction*, 1–19.
<https://doi.org/10.1080/10447318.2022.2121219>
- Deci, E. L., & Ryan, R. M. (2000). The „What“ and „Why“ of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From Game Design Elements to Gamefulness: Defining „Gamification“. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 9–15.
<https://doi.org/10.1145/2181037.2181040>

- DeVito, J. A. (2000). *Human Communication: The Basic Course*. Longman.
- Dichev, C., & Dicheva, D. (2017). Gamifying Education: What Is Known, What Is Believed and What Remains Uncertain: A Critical Review. *International Journal of Educational Technology in Higher Education*, 14(1), 9. <https://doi.org/10.1186/s41239-017-0042-5>
- 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.
- Edo, X. P., & Lucio, V. G. G. (2021). *The Alexa Revolution: Alexa Skills Development*. Xavier Portilla Edo.
- Fencott, P. C. (Hrsg.). (2012). *Game Invaders: The Theory and Understanding of Computer Games* [Electronic resource]. IEEE Press ; Wiley.
- Fish, M. T., & Saul, A. D. (2019). The Gamification of Meditation: A Randomized-Controlled Study of a Prescribed Mobile Mindfulness Meditation Application in Reducing College Students' Depression. *Simulation & Gaming*, 50(4), 419–435. <https://doi.org/10.1177/1046878119851821>
- Fortes Tondello, G. (2019). *Dynamic Personalization of Gameful Interactive Systems* [Doctoral Thesis, University of Waterloo]. <https://uwspace.uwaterloo.ca/handle/10012/14807>
- Friberg, J., & Gärdenfors, D. (2004). Audio Games: New Perspectives on Game Audio. *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology - ACE '04*, 148–154. <https://doi.org/10.1145/1067343.1067361>
- Fujishin, R. (2019). *Natural Bridges in Interpersonal Communication*. Routledge.
- Garcia, F. E., de Almeida Neris, V. P., Garcia, F. E., & de Almeida Neris, V. P. (2013). Design Guidelines for Audio Games. In M. Kurosu & M. Kurosu (Hrsg.),

Proceedings of the 15th International Conference on Human-Computer Interaction (Bd. 8005, S. 229–238). Springer. https://doi.org/10.1007/978-3-642-39262-7_26

Groening, C., & Binnewies, C. (2019). “Achievement Unlocked!”—The Impact of Digital Achievements as a Gamification Element on Motivation and Performance. *Computers in Human Behavior*, *97*, 151–166. <https://doi.org/10.1016/j.chb.2019.02.026>

Guzman, A. L. (2019). Voices in and of the Machine: Source Orientation Toward Mobile Virtual Assistants. *Computers in Human Behavior*, *90*, 343–350. <https://doi.org/10.1016/j.chb.2018.08.009>

Hakulinen, L., Auvinen, T., & Korhonen, A. (2015). The Effect of Achievement Badges on Students’ Behavior: An Empirical Study in a University-Level Computer Science Course. *International Journal of Emerging Technologies in Learning (iJET)*, *10*(1), 18. <https://doi.org/10.3991/ijet.v10i1.4221>

Hamari, J. (2017). Do Badges Increase User Activity? A Field Experiment on the Effects of Gamification. *Computers in Human Behavior*, *71*, 469–478. <https://doi.org/10.1016/j.chb.2015.03.036>

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS)*, 3025–3034. <https://doi.org/10.1109/HICSS.2014.377>

Host a Custom Skill as a Web Service | Alexa Skills Kit. (2022, März 29). Amazon Alexa. <https://developer.amazon.com/en-US/docs/alexa/custom-skills/host-a-custom-skill-as-a-web-service.html>

- Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, 37(1), 81–88. <https://doi.org/10.1080/02763869.2018.1404391>
- Hunicke, R., LeBlanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. *Proceedings of the AAAI Workshop on Challenges in Game AI*, 4, 1–5.
- Huotari, K., & Hamari, J. (2017). A Definition for Gamification: Anchoring Gamification in the Service Marketing Literature. *Electronic Markets*, 27(1), 21–31. <https://doi.org/10.1007/s12525-015-0212-z>
- Jørgensen, K. (2008). Left in the Dark: Playing Computer Games with the Sound Turned Off. In *From Pac-Man to Pop Music*. Routledge.
- Jørgensen, K. (2006). On the Functional Aspects of Computer Game Audio. *Proceedings of the Audio Mostly Conference 2006*. <https://bora.uib.no/bora-xmlui/handle/1956/6734>
- Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. Pfeiffer.
- Karoly, P. (1993). Mechanisms of Self-Regulation: A Systems View. *Annual Review of Psychology*, 44(1), 23–52. <https://doi.org/10.1146/annurev.ps.44.020193.000323>
- Klock, A. C. T., Gasparini, I., Pimenta, M. S., & ... (2020). Tailored gamification: A review of literature. In *International Journal of ...*. Elsevier. <https://www.sciencedirect.com/science/article/pii/S1071581920300975>
- Koivisto, J., & Hamari, J. (2019). The Rise of Motivational Information Systems: A Review of Gamification Research. *International Journal of Information Management*, 45, 191–210. <https://doi.org/10.1016/j.ijinfomgt.2018.10.013>
- Krath, J., Schürmann, L., & von Korfflesch, H. F. O. (2021). Revealing the Theoretical Basis of Gamification: A Systematic Review and Analysis of Theory in Research

- on Gamification, Serious Games and Game-Based Learning. *Computers in Human Behavior*, 125, 106963. <https://doi.org/10.1016/j.chb.2021.106963>
- La Guardia, J. G., Ryan, R. M., Couchman, C. E., & Deci, E. L. (2000). Within-Person Variation in Security of Attachment: A Self-Determination Theory Perspective on Attachment, Need Fulfillment, and Well-Being. *Journal of Personality and Social Psychology*, 79(3), 367–384. <https://doi.org/10.1037/0022-3514.79.3.367>
- 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. <https://doi.org/10.1177/1046878118774385>
- 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. <https://doi.org/10.1016/j.chb.2015.08.008>
- Laricchia, F. (2023, März). *Smart Speaker Market Revenue from 2015 to 2027*. Statista. <https://www.statista.com/forecasts/1367973/smart-speaker-market-revenue-worldwide>
- Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–31. <https://doi.org/10.1145/3274371>
- Lee, D. (2023, März 4). Amazon’s big dreams for Alexa fall short. *Financial Times*.
- Lee, D., Sah, Y. J., & Lee, S. (2019). Improving Usability Perception of Error-Prone AI Speakers: Elaborated Feedback Mitigates Negative Consequences of Errors. *International Journal of Human-Computer Interaction*, 35(17), 1645–1652. <https://doi.org/10.1080/10447318.2018.1561069>
- Lima, M. R., Su, T., Jouaiti, M., Wairagkar, M., Malhotra, P., Soreq, E., Barnaghi, P., & Vaidyanathan, R. (2023). Discovering Behavioral Patterns Using Conversational

- Technology for In-Home Health and Well-Being Monitoring. *IEEE Internet of Things Journal*, 10(21), 18537–18552.
<https://doi.org/10.1109/JIOT.2023.3290833>
- Ling, E. C., Tussyadiah, I., Tuomi, A., Stienmetz, J., & Ioannou, A. (2021). Factors Influencing Users' Adoption and Use of Conversational Agents: A Systematic Review. *Psychology & Marketing*, 38(7), 1031–1051.
<https://doi.org/10.1002/mar.21491>
- Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, 3(2), 157–189. [https://doi.org/10.1016/0030-5073\(68\)90004-4](https://doi.org/10.1016/0030-5073(68)90004-4)
- Locke, E. A., & Latham, G. P. (2002). Building a Practically Useful Theory of Goal Setting and Task Motivation. *American Psychologist*, 57(9), 705–717.
- Lopatovska, I., Rink, K., Knight, I., Raines, K., Cosenza, K., Williams, H., Sorsche, P., Hirsch, D., Li, Q., & Martinez, A. (2019). Talk to Me: Exploring User Interactions with the Amazon Alexa. *Journal of Librarianship and Information Science*, 51(4), 984–997. <https://doi.org/10.1177/0961000618759414>
- Luger, E., & Sellen, A. (2016). „Like Having a Really Bad PA“: The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5286–5297.
<https://doi.org/10.1145/2858036.2858288>
- Marczewski, A. (2018). *Even Ninja Monkeys like to play: Gamification, game thinking and motivational design* (Unicorn edition). Gamified UK.
- Mari, A. (2019). *Voice Commerce: Understanding Shopping-Related Voice Assistants and their Effect on Brands*. <https://doi.org/10.5167/UZH-197725>

- Mazarakis, A. (2015). Using Gamification for Technology Enhanced Learning: The Case of Feedback Mechanisms. *Bulletin of the IEEE Technical Committee on Learning Technology*, 4(17), 6–9.
- Mazarakis, A. (2017). Gamification: Eine experimentelle Untersuchung der Spielelemente Abzeichen und Story. *Mensch und Computer 2017 - Tagungsband: Spielend einfach interagieren*, 3–14. <https://doi.org/10.18420/MUC2017-MCI-0188>
- Mazarakis, A. (2021). Gamification Reloaded: Current and Future Trends in Gamification Science. *I-Com*, 20(3), 279–294. <https://doi.org/10.1515/icom-2021-0025>
- Mazarakis, A., & Bräuer, P. (2020). Gamification of an Open Access Quiz with Badges and Progress Bars: An Experimental Study with Scientists. In J. Koivisto, M. Bujic, & J. Hamari (Hrsg.), *Proceedings of the 4th International GamiFIN Conference—GamiFIN 2020* (Bd. 2637, S. 62–71). CEUR-WS.org.
- Mazarakis, A., & Bräuer, P. (2022). Gamification und die Wahrnehmung von Punkten – Eine Umfragestudie. *MuC'22 Workshops*. <https://doi.org/10.18420/muc2022-mci-ws08-417>
- McTear, M. (2017). The Rise of the Conversational Interface: A New Kid on the Block? In J. F. Quesada, F.-J. Martín Mateos, & T. López Soto (Hrsg.), *Future and Emerging Trends in Language Technology. Machine Learning and Big Data* (Bd. 10341, S. 38–49). Springer International Publishing. https://doi.org/10.1007/978-3-319-69365-1_3
- McTear, M. (2021). *Conversational Ai: Dialogue Systems, Conversational Agents, and Chatbots*. Morgan & Claypool Publishers.
- McTear, M., Callejas, Z., & Griol, D. (2016). *The Conversational Interface*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-32967-3>

- Mekler, E. D., Bruehlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards Understanding the Effects of Individual Gamification Elements on Intrinsic Motivation and Performance. *Computers in Human Behavior*, 71, 525–534. <https://doi.org/10.1016/j.chb.2015.08.048>
- Monteiro, P., Goncalves, G., Coelho, H., Melo, M., & Bessa, M. (2021). Hands-Free Interaction in Immersive Virtual Reality: A Systematic Review. *IEEE Transactions on Visualization and Computer Graphics*, 27(5), 2702–2713. <https://doi.org/10.1109/TVCG.2021.3067687>
- Mora, A., Riera, D., Gonzalez, C., & Arnedo-Moreno, J. (2015). A Literature Review of Gamification Design Frameworks. *2015 7th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games)*, 1–8. <https://doi.org/10.1109/VS-GAMES.2015.7295760>
- Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. <https://doi.org/10.1109/MRA.2012.2192811>
- Morschheuser, B., Hamari, J., Werder, K., & Abe, J. (2017). *How to Gamify? A Method For Designing Gamification*. Hawaii International Conference on System Sciences. <https://doi.org/10.24251/HICSS.2017.155>
- Murad, C., & Munteanu, C. (2020). Designing Voice Interfaces: Back to the (Curriculum) Basics. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3313831.3376522>
- Murad, C., Munteanu, C., Clark, L., & Cowan, B. R. (2018). Design Guidelines for Hands-Free Speech Interaction. *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '18*, 269–276. <https://doi.org/10.1145/3236112.3236149>

- Murad, C., Munteanu, C., R. Cowan, B., & Clark, L. (2021). Finding a New Voice: Transitioning Designers from GUI to VUI Design. *CUI 2021 - 3rd Conference on Conversational User Interfaces*, 1–12. <https://doi.org/10.1145/3469595.3469617>
- Myers, C., Furqan, A., Nebolsky, J., Caro, K., & Zhu, J. (2018). Patterns for How Users Overcome Obstacles in Voice User Interfaces. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–7. <https://doi.org/10.1145/3173574.3173580>
- Nacke, L. E., & Deterding, S. (2017). The Maturing of Gamification Research. *Computers in Human Behavior*, 71, 450–454. <https://doi.org/10.1016/j.chb.2016.11.062>
- Oliveira, W., Pastushenko, O., Rodrigues, L., Toda, A. M., Palomino, P. T., Hamari, J., & Isotani, S. (2021). *Does Gamification Affect Flow Experience? A Systematic Literature Review* (arXiv:2106.09942). arXiv. <https://doi.org/10.48550/arXiv.2106.09942>
- Palmquist, A. (2021). Lost in Translation: A Study of (mis)conceptions, (mis)communication and Concerns When Implementing Gamification in Corporate (re)training. *Proceedings of the 54th Hawaii International Conference on System Sciences*, 1375–1384. <https://doi.org/10.24251/HICSS.2021.166>
- Pearl, C. (2016). *Designing Voice User Interfaces: Principles of Conversational Experiences*. O'Reilly Media, Inc.
- Pelling, N. (2011, August 9). The (short) prehistory of “gamification”. *Funding Startups (& Other Impossibilities)*. <https://nanodome.wordpress.com/2011/08/09/the-short-prehistory-of-gamification/>
- Pereira, G., Brisson, A., Prada, R., Paiva, A., Bellotti, F., Kravcik, M., & Klamka, R. (2012). Serious Games for Personal and Social Learning & Ethics: Status and

Trends. *Procedia Computer Science*, 15, 53–65.

<https://doi.org/10.1016/j.procs.2012.10.058>

Perez Garcia, M., & Saffon Lopez, S. (2020). Exploring the Uncanny Valley Theory in the Constructs of a Virtual Assistant Personality. In Y. Bi, R. Bhatia, & S. Kapoor (Hrsg.), *Intelligent Systems and Applications* (Bd. 1037, S. 1017–1033). Springer International Publishing. https://doi.org/10.1007/978-3-030-29516-5_76

Planner, A. (2018, Juli 5). *How Many Users Should You Expect for Your Voice Skill?* <https://www.screenmedia.co.uk/news/how-many-users-should-you-expect-for-your-voice-skill/>

Pradhan, A., Mehta, K., & Findlater, L. (2018). „Accessibility Came by Accident“: Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3173574.3174033>

Putz, L.-M., Hofbauer, F., & Treiblmaier, H. (2020). Can Gamification Help to Improve Education? Findings from a Longitudinal Study. *Computers in Human Behavior*, 110, 106392. <https://doi.org/10.1016/j.chb.2020.106392>

Rackwitz, R., & Niesenhaus, J. (2018). #3 *Gamification für Voice Alexa Skill* (3). Abgerufen 9. Februar 2023, von <https://open.spotify.com/episode/3lWthrRE81BvYwCaxB4nsi>

Rapp, A., Hopfgartner, F., Hamari, J., Linehan, C., & Cena, F. (2019). Strengthening Gamification Studies: Current Trends and Future Opportunities of Gamification Research. *International Journal of Human-Computer Studies*, 127, 1–6. <https://doi.org/10.1016/j.ijhcs.2018.11.007>

Reinartz, W. J. (2010). Understanding Customer Loyalty Programs. In M. Krafft & M. K. Mantrala (Hrsg.), *Retailing in the 21st Century* (S. 409–427). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72003-4_25

- Rhee, C. E., & Choi, J. (2020). Effects of Personalization and Social Role in Voice Shopping: An Experimental Study on Product Recommendation by a Conversational Voice Agent. *Computers in Human Behavior, 109*, 106359. <https://doi.org/10.1016/j.chb.2020.106359>
- Röber, N., & Masuch, M. (2005). *Leaving the Screen New Perspectives in Audio-Only Gaming*. <https://smartech.gatech.edu/handle/1853/50168>
- Rodrigues, L., Toda, A. M., Palomino, P. T., Oliveira, W., & Isotani, S. (2020). Personalized Gamification: A Literature Review of Outcomes, Experiments, and Approaches. *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 699–706. <https://doi.org/10.1145/3434780.3436665>
- Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., & Landay, J. A. (2018). Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1*(4), 1–23. <https://doi.org/10.1145/3161187>
- Rzepka, C., Berger, B., & Hess, T. (2020). Why Another Customer Channel? Consumers' Perceived Benefits and Costs of Voice Commerce. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 4079–4088. <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/fc16e5de-a705-48c0-b1a5-9888525e2966/content>
- Rzepka, C., Berger, B., & Hess, T. (2022). Voice Assistant vs. Chatbot – Examining the Fit Between Conversational Agents' Interaction Modalities and Information Search Tasks. *Information Systems Frontiers, 24*(3), 839–856. <https://doi.org/10.1007/s10796-021-10226-5>
- Sailer, M., Hense, J., Mandl, H., & Klevers, M. (2013). Psychological Perspectives on Motivation through Gamification. *IxD&A, 19*, 28–37.

- 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. <https://doi.org/10.1016/j.chb.2016.12.033>
- Sailer, M., & Homner, L. (2020). The Gamification of Learning: A Meta-Analysis. *Educational Psychology Review*, *32*(1), 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Sakr, M. (2021). ‘She’s Nice Company and a Good Friend’: Thinking with Haraway to Reconceptualise Children’s Playful Interactions with Alexa in the Family Home. In J. Mikats, S. Kink-Hampersberger, & L. Oates-Indruchová (Hrsg.), *Creative Families: Gender and Technologies of Everyday Life* (S. 167–188). Springer International Publishing. https://doi.org/10.1007/978-3-030-70803-0_8
- Saleem, A. N., Noori, N. M., & Ozdamli, F. (2022). Gamification Applications in E-learning: A Literature Review. In *TECHNOLOGY KNOWLEDGE AND LEARNING* (Bd. 27, Nummer 1, S. 139–159). <https://doi.org/10.1007/s10758-020-09487-x>
- Seaborn, K., & Fels, D. I. (2015). Gamification in Theory and Action: A Survey. *International Journal of Human-Computer Studies*, *74*, 14–31. <https://doi.org/10.1016/j.ijhcs.2014.09.006>
- Shahri, A., Hosseini, M., Phalp, K., Taylor, J., & Ali, R. (2014). Towards a Code of Ethics for Gamification at Enterprise. In U. Frank, P. Loucopoulos, Ó. Pastor, & I. Petrounias (Hrsg.), *The Practice of Enterprise Modeling* (Bd. 197, S. 235–245). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45501-2_17
- Shortt, M., Tilak, S., Kuznetcova, I., Martens, B., & Akinkuolie, B. (2023). Gamification in Mobile-Assisted Language Learning: A Systematic Review of Duolingo Literature from Public Release of 2012 to Early 2020. *Computer Assisted*

Language Learning, 36(3), 517–554.

<https://doi.org/10.1080/09588221.2021.1933540>

Spil, T., Sunyaev, A., Thiebes, S., & Van Baalen, R. (2017). *The Adoption of Wearables for a Healthy Lifestyle: Can Gamification Help?* Hawaii International Conference on System Sciences. <https://doi.org/10.24251/HICSS.2017.437>

Stachyra, G. (2012). The Radio Plays Games. *European Journal of Communication*, 27(3), 241–256. <https://doi.org/10.1177/0267323112452832>

Susi, T., Johannesson, M., & Backlund, P. (2007). *Serious Games – An Overview*.

Thomas, N. J., Baral, R., & Crocco, O. S. (2022). Gamification for HRD: Systematic Review and Future Research Directions. *Human Resource Development Review*, 21(2), 198–224. <https://doi.org/10.1177/15344843221074859>

Thorpe, A. S., & Roper, S. (2019). The Ethics of Gamification in a Marketing Context. *Journal of Business Ethics*, 155(2), 597–609. <https://doi.org/10.1007/s10551-017-3501-y>

Tobar, H. F. (2017). *Are game mechanics game design elements?* Quora. <https://www.quora.com/Are-game-mechanics-game-design-elements>

Toda, A. M., Klock, A. C. T., Oliveira, W., Palomino, P. T., Rodrigues, L., Shi, L., Bittencourt, I., Gasparini, I., Isotani, S., & Cristea, A. I. (2019). Analysing Gamification Elements in Educational Environments Using an Existing Gamification Taxonomy. *Smart Learning Environments*, 6(1), 16. <https://doi.org/10.1186/s40561-019-0106-1>

Urbanek, M., & Güldenpfennig, F. (2019). Celebrating 20 Years of Computer-Based Audio Gaming. *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, 90–97. <https://doi.org/10.1145/3356590.3356605>

- Vasudevan, K., & Chan, N. K. (2022). Gamification and Work Games: Examining Consent and Resistance Among Uber Drivers. *New Media & Society*, 24(4), 866–886. <https://doi.org/10.1177/14614448221079028>
- Vooris, R., Blaszk, M., & Purrington, S. (2019). Understanding the Wearable Fitness Tracker Revolution. *International Journal of the Sociology of Leisure*, 2(4), 421–437. <https://doi.org/10.1007/s41978-018-00022-y>
- Wang, J., Li, Y., Yang, S., Dong, S., & Li, J. (2023). Waiting Experience: Optimization of Feedback Mechanism of Voice User Interfaces Based on Time Perception. *IEEE Access*, 11, 21241–21251. <https://doi.org/10.1109/ACCESS.2023.3250278>
- Warmelink, H., Koivisto, J., Mayer, I., Vesa, M., & Hamari, J. (2018). Gamification of the Work Floor: A Literature Review of Gamifying Production and Logistics Operations. *Proceedings of the 51st Hawaii International Conference on System Sciences*, 1108–1117.
- Warmelink, H., Koivisto, J., Mayer, I., Vesa, M., & Hamari, J. (2020). Gamification of Production and Logistics Operations: Status Quo and Future Directions. *Journal of Business Research*, 106, 331–340. <https://doi.org/10.1016/j.jbusres.2018.09.011>
- Werbach, K., & Hunter, D. (2012). *For the Win: How Game Thinking Can Revolutionize Your Business*. Wharton Digital Press.
- What Is Alexa Voice ID?* (o. J.). Abgerufen 15. September 2023, von https://www.amazon.com/-/de/gp/help/customer/display.html?ref_=hp_left_v4_sib&nodeId=GYCXKY2AB2QWZT2X
- Windasari, N. A., & Lin, F. (2021). Why Do People Continue Using Fitness Wearables? The Effect of Interactivity and Gamification. *SAGE Open*, 11(4), 21582440211056610. <https://doi.org/10.1177/21582440211056606>

- Winkler, R., Söllner, M., Neuweiler, M. L., Conti Rossini, F., & Leimeister, J. M. (2019). Alexa, Can You Help Us Solve This Problem?: How Conversations With Smart Personal Assistant Tutors Increase Task Group Outcomes. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19*, 1–6. <https://doi.org/10.1145/3290607.3313090>
- Wunderlich, N. V., Gustafsson, A., Hamari, J., Parvinen, P., & Haff, A. (2020). The Great Game of Business: Advancing Knowledge on Gamification in Business Contexts. *Journal of Business Research*, 106, 273–276. <https://doi.org/10.1016/j.jbusres.2019.10.062>
- Xu, H., & Hamari, J. (2022). How to Improve Creativity: A Study of Gamification, Money, and Punishment. *Behaviour & Information Technology*, 1–15. <https://doi.org/10.1080/0144929X.2022.2133634>
- Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The Impact of Gamification on Learning and Instruction: A Systematic Review of Empirical Evidence. *Educational Research Review*, 30, 100326. <https://doi.org/10.1016/j.edurev.2020.100326>
- Zichermann, G., & Cunningham, C. (2011). *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. O'Reilly Media.

Appendix

Workshare of Co-authored Publications

Statutory Declaration

Curriculum vitae

Workshare of Co-authored Publications

Details on the doctoral candidate's workshare regarding co-authored manuscripts included in this dissertation.

Study 1

Authors: Athanasios Mazarakis & Paula Bräuer

Candidate's estimated workshare:

| Conceptualization | Planning | Execution | Manuscript Preparation |
|--------------------------|-----------------|------------------|-----------------------------------|
| 30% | 60% | 80% | 60% |

Study 2

Authors: Paula Bräuer & Athanasios Mazarakis

Candidate's estimated workshare:

| Conceptualization | Planning | Execution | Manuscript Preparation |
|--------------------------|-----------------|------------------|-----------------------------------|
| 80% | 60% | 10% | 80% |

Study 3

Authors: Paula Bräuer & Athanasios Mazarakis

Candidate's estimated workshare:

| Conceptualization | Planning | Execution | Manuscript Preparation |
|--------------------------|-----------------|------------------|-----------------------------------|
| 90% | 70% | 60% | 80% |

Study 4

Authors: Paula Bräuer , Florian Werner & Athanasios Mazarakis

Candidate's estimated workshare:

| Conceptualization | Planning | Execution | Manuscript Preparation |
|--------------------------|-----------------|------------------|-----------------------------------|
| 80% | 70% | 10% | 90% |

Place, Date

Signature of the doctoral candidate

Place, Date

Signature of the supervisor

Statutory Declaration

I hereby confirm that - apart from my supervisors' guidance - the content and design of this thesis is my own work, and all used sources have been listed and duly acknowledged in the text. The thesis has not been submitted for any other degree or professional qualification. Parts based on jointly-authored publications have been clearly denoted as such, and my own respective workshare documented. The thesis has been prepared subject to the Rules of Good Scientific Practice of the German Research Foundation. I confirm that no academic degree has ever been withdrawn from me.

Place, Date

Signature of the doctoral candidate

Paula Bräuer

E-Mail: paula.braeuer@live.de
Nationality: German
ORCID ID: 0000-0001-5903-8829
GitHub: <https://github.com/pbrzbw>

Professional Experience

Since 2020 Research assistant/PhD candidate at Kiel University (TRANSES project)

2018 - 2020 Research assistant at Kiel University (OA-FWM project)

2015 - 2018 Student assistant at Kiel University

Higher Education

2015 – 2018 Master of Science in Business Information Technology at Kiel University (6 Semester)

2011 – 2015 Bachelor of Science in Business Information Technology at Kiel University (8 Semester)

Teaching

21.06.22 & 20.06.23 TRANSENS Lecture series – “Information design for transdisciplinary work“

winter term 21/22 Masterseminar - Incentives for Collective Intelligence and User Generated Content

winter term 20/21 Masterseminar - Incentives for Collective Intelligence and User Generated Content

winter term 19/20 Masterseminar - Incentives for Collective Intelligence and User Generated Content

winter term 18/19 Masterseminar - Incentives for Collective Intelligence and User Generated Content

Publications

JOURNALS

Mazarakis, A., & Bräuer, P. (2023). Gamification is Working, but Which One Exactly? Results from an Experiment with Four Game Design Elements. *International Journal of Human-Computer Interaction*, 39(3), 612–627. <https://doi.org/10.1080/10447318.2022.2041909>

Bräuer, P., & Mazarakis, A. (2022). “Alexa, Can We Design Gamification Without a Screen?”—Implementing Cooperative and Competitive Audio-Gamification for Intelligent Virtual Assistants. *Computers in Human Behavior*, 135, 107362. <https://doi.org/10.1016/j.chb.2022.107362>

Bräuer, P., & Mazarakis, A. (2022). How to Design Audio-Gamification for Language Learning with Amazon Alexa?—A Long-Term Field Experiment. *International Journal of Human-Computer Interaction*. <https://doi.org/10.1080/10447318.2022.2160228>

PAPERS IN CONFERENCE PROCEEDINGS (PEER REVIEWED)

Bräuer, P., Berg, M., Mazarakis, A., & Peters, I. (2023). Movement in Virtual Time: How Virtual Reality Can Support Long-Term Thinking. *Mensch Und Computer* 2023, 477–481. <https://doi.org/10.1145/3603555.3608569>

Bräuer, P., Schwarz, L., & Peters, I. (2023). Transdisziplinäre Entwicklung eines WebGIS für die Atommüll-Endlagerstandortsuche in Deutschland. *INFORMATIK 2023 - Designing Futures: Zukünfte gestalten*, 777–782. https://doi.org/10.18420/INF2023_87

Bräuer, P., Werner, F., & Mazarakis, A. (2023). Extracting Game Design Elements from Voice-Enabled Games: A Review of Amazon Alexa Skills. *Proceedings of the 7th International GamiFIN Conference*, 47–56.

Schwarz, L., Bräuer, P., Consten, W., Schöler, M., Guth, A., Nissen, N., & Gaebel, B. (2023). Ein partizipatives WebGIS für die Endlagerstandortsuche: Eine transdisziplinäre Studie zu einer inklusiven Kommunikationsplattform [TRANSENS Bericht].

Bräuer, P., & Mazarakis, A. (2022). A Speech-Based AI for Political Participation. *Mensch Und Computer* 2022, 462–466. <https://doi.org/10.1145/3543758.3549889>

Mazarakis, A., & Bräuer, P. (2022). Gamification und die Wahrnehmung von Punkten – Eine Umfragestudie. *MuC'22 Workshops*. <https://doi.org/10.18420/muc2022-mci-ws08-417>

Schwarz, L., & Bräuer, P. (2022). An Exploratory PPGIS for the Nuclear Waste Repository Siting Procedure in Germany – a Transdisciplinary Approach to Enable Meaningful Participation? *GI_Forum 2022, Volume 10*, 77–90. https://doi.org/10.1553/giscience2022_01_s77

Bräuer, P., Jacob, A., Mazarakis, A., & Peters, I. (2021). Vertrauenswürdigkeit und Wissenschaftlichkeit von Visualisierungen im partizipativen Prozess der Endlagersuche für hoch radioaktive Abfälle in Deutschland. *Gemeinschaften in Neuen Medien. Digitale Partizipation in Hybriden Realitäten Und Gemeinschaften.*, 302–312. <https://doi.org/10.25368/2022.52>

Fraser, N., Bräuer, P., & Peters, I. (2021). Altmetrics for evaluation of medical research in Germany. *Proceedings of 18th International Conference on Scientometrics & Informetrics*.

Lemke, Steffen, Bräuer, Paula, & Peters, Isabella. (2021). Does the General Public Share Research on Twitter? A Case Study on the Online Conversation About the Search for a Nuclear Repository in Germany. *Universität Regensburg*. <https://doi.org/10.5283/EPUB.44940>

Schwarz, L., & Bräuer, P. (2021). Improving Participation in the German Search for a Nuclear Waste Repository Site: A WebGIS as a Transdisciplinary Approach to Support Dialogue? *Safety of Nuclear Waste Disposal*, 1, 209–210. <https://doi.org/10.5194/sand-1-209-2021>

Sierra, R., Mbah, M., Schwarz, L., Themann, D., Beninghaus, C., Becker, F., & Bräuer, P. (2021). Experiences and Challenges of Digital Transdisciplinary Formats (td-Formats) in Complex and Contested Research Fields. *ITD 21 - International Transdisciplinary Conference*.

Bräuer, P., & Mazarakis, A. (2020). Visualization of Turnover Rate in a Warehouse Using Augmented Reality: A Demo with the Microsoft HoloLens. *Proceedings of Mensch Und Computer 2020*, 519–522. <https://doi.org/10.1145/3404983.3410422>

Mazarakis, A., & Bräuer, P. (2020). Gamification of an Open Access Quiz with Badges and Progress Bars: An Experimental Study with Scientists. In J. Koivisto, M. Bujic, & J. Hamari (Eds.), *Proceedings of the 4th International GamiFIN Conference—GamiFIN 2020* (Vol. 2637, pp. 62–71). CEUR-WS.org.

Bräuer, P., & Mazarakis, A. (2019a). Badges or a Leaderboard? How to Gamify an Augmented Reality Warehouse Setting. In J. Koivisto & Hamari (Eds.), *Proceedings of the 3rd International GamiFIN Conference—GamiFIN 2019* (Vol. 2359, pp. 229–240). CEUR-WS.org.

Bräuer, P., & Mazarakis, A. (2019b). Erhöhung der Motivation für Open Access durch Gamification. *MuC'19 Workshops*, 43–48. <https://doi.org/10.18420/muc2019-ws-564>

Bräuer, P., & Mazarakis, A. (2018a). AR in Order-Picking – Experimental Evidence with Microsoft HoloLens. *Mensch und Computer 2018 – Workshopband*, Bonn: Gesellschaft für Informatik e.V., 361–368. <https://doi.org/10.18420/MUC2018-WS07-0463>

Bräuer, P., & Mazarakis, A. (2018b). Gamification und Augmented Reality für Lagerprozesse. *Mensch und Computer 2018 - Workshopband*. Bonn: Gesellschaft für Informatik e.V., 103–111 <https://doi.org/10.18420/MUC2018-WS03-0476>

Mazarakis, A., & Bräuer, P. (2017). Welche Gamification motiviert? Ein Experiment zu Abzeichen, Feedback, Fortschrittsanzeige und Story. *GeNeMe '17 - Gemeinschaften in Neuen Medien*, 20. Tagung, 246–255.