Bräuer, Paula; Werner, Florian; Mazarakis, Athanasios

Conference Paper — Published Version

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

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

Paula Bräuer [1], Florian Werner [1] and Athanasios Mazarakis [2]

[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 thee-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

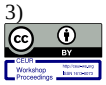
1. Introduction

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)



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" category[2]. In contrast, an examination of the "games and quizzes" category in the Amazon Alexa Store suggested more than five hundred times the number of skills[3].

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.

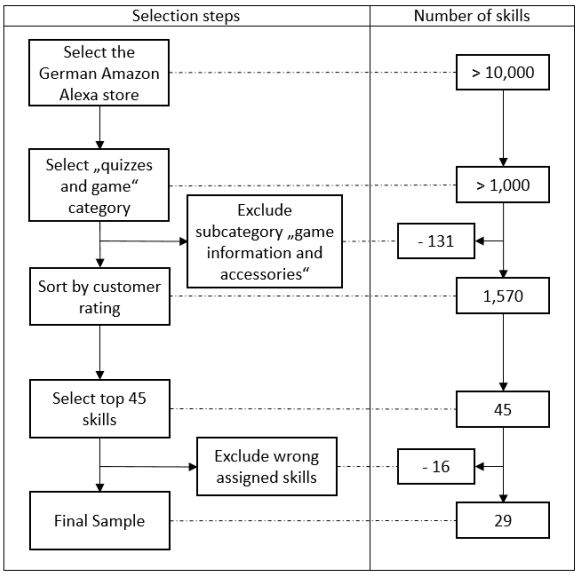


Figure 1: Flowchart of the Alexa skill selection

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

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

[3] https://www.amazon.de/s?bbn=10068461031&rh=n%3A10068460031%2Cn%3A10536644031&dc&qid=1672997334&rnid=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 Ciccío 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.

N*arrative 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.

	Rating		Game type by Amazon		Game type	
Name of the skill	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