Lehren und Lernen mit Intelligenten Systemen

Bachelorarbeit

Gender Differences in the Impact of Gamification Elements on Performance and Anxiety

Robin Gebert

Studiengang: Informatik - B.A. (Lehramt)

Prüfer/in: Jun.-Prof. Dr. rer. nat. Maria Wirzberger

Betreuer/in: Nadine Nicole Koch, M.Sc.
Niklas Meißner, M.Sc.

Beginn am: 16. Mai

Beendet am: 16. September 2023

Abstract

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Keywords: Gamification, Gender

Gender Differences in the Impact of Gamification Elements on Performance and Anxiety

Introduction

The often mandated transition to remote education during the COVID-19 pandemic highlighted the indispensable role of digital learning environments (DLEs) (García-Morales et al., 2021). The usage of digital learning resources saw a significant increase during this period; in 2019, 54% of students utilized such resources, with the figure rising to 70% in 2020. In Germany, the percentage of scholars aged ten to fifteen engaging with digital learning materials doubled from 32% to 64% within the same timeframe (des statistischen Bundesamts, 2020).

The integration of gamified elements into DLEs is a common practice, aimed at enhancing motivation and enjoyment (González et al., 2014; Jackson & McNamara, 2013). However, despite the positive impacts, the introduction of some elements can also lead to negative outcomes, including demotivation (Almeida et al., 2021). The concept of gamification has attracted substantial interest within the educational sciences, becoming a prevalent topic (Swacha, 2021). In the realm of computer science, the application of gamified elements is well-documented, demonstrating a important presence due to the inherent integration of technology in the field (Dichev & Dicheva, 2017).

Although there is extensive research on the general application of gamification, the effects related to individual factors, such as gender, are less understood and warrant further investigation (Dehghanzadeh et al., 2024; Oliveira et al., 2023).

Digital learning environments, especially Tutorial systems offer a wide range of features and flexibility to support students learning processes. These systems can range from simple instructive texts to simulations and virtual realities, serving as models that simplify aspects of the real world to reduce complexity mostly from interconnection and context of knowledge for both the machine and the user (Psotka et al., 1988). Those environments are in use at higher education institutions and since the COVID-19 pandemic, they have become more important (El Hadbi et al., 2024) and widely researched, especially in the field of computer science (Zawacki-Richter et al., 2019).

Tutorial systems are often enhanced with some sort of intelligence (ITS). A dynamic adaptation to learner's needs, incorporating usage factors such as performance, but also external factors such as age, culture or gender (González et al., 2014; Nkambou et al., 2010). The role of Artificial Intelligence in this process could also become more important in the future, as it could have a significant, yet unknown impact on ITS (Zawacki-Richter et al., 2019).

Tutoring Systems are often used in combination with gamified elements, significantly improving the learning experience (Dermeval et al., 2019). Jackson and McNamara (2013) found that the use of gamified elements in tutoring systems significantly improved the motivation and performance of students compared to traditional tutoring systems which lead to boredom and disengagement after long periods of use.

Incorporating gamified elements not only enhances the engagement and motivation within the ITS but also necessitates mechanisms for tracking progress, such as content unlocking (González et al., 2014). The evolving landscape of ITS research also includes emotional and relational dynamics, linking student emotions and teacher-student relationships to learning efficacy and motivation (Woolf et al., 2010). These insights have led to the development of digital companions, often named pedagogical agents, within ITS that significantly boost the learning potential and self-concept of students, particularly those who are low-achieving. Intriguingly, a study noted that ITS programs with a male companion were muted twice as often as those with a female companion, highlighting potential gender differences that could be explored to enhance the predictive capabilities of the ITS (Woolf et al., 2010).

Gender, as a concept within social sciences, refers to more than the binary categorization of male and female. It encompasses a range of identities and experiences that are shaped by a complex interplay of biological, psychological, and social factors. Gender is not solely determined by biological characteristics; instead, it is increasingly recognized as a spectrum, acknowledging the presence of diverse gender identities beyond the traditional binary understanding (Lindqvist et al., 2021). Socialization plays a critical role in shaping gender identity. It influences how individuals perceive themselves and interact with their surroundings based on the gender norms prevalent within their society. These norms dictate behaviours, roles, and expectations, which are often internalized from an early age through various socialization agents like family, media, educational institutions, and peer groups (Kampshoff & Wiepcke, 2012). While acknowledging the spectrum of gender identities, this thesis will focus primarily on the binary categorization of gender—male and female. This approach does not negate the validity of non-binary or genderqueer identities but rather limits the scope of investigation to traditional gender roles within the binary framework.

Gender is a critical variable to consider when deploying tutorial systems due to its significant influence on learning preferences and outcomes. Research indicates that male and female students exhibit distinct preferences for learning modalities and react differently to adaptive learning technologies. For instance, studies have shown that while male students often prefer multimodal instructional approaches, female students tend to favour single-mode learning, particularly kinesthetic styles (Wehrwein et al., 2007). Additionally, the use of adaptive learning technologies has demonstrated a more pronounced improvement in performance among male students compared to their female counterparts in subjects like Mathematics and Portuguese (De Santana et al., 2016). Moreover, technological proficiency and communication challenges in online settings have been identified as factors affecting learning satisfaction, with older students showing a stronger preference for face-to-face learning, which may be due to less familiarity with digital technologies (Dabaj, 2009). Recognizing and addressing gender differences in learning styles and preferences is essential for tailoring educational technologies and strategies, thereby optimizing tutorial systems to enhance learning efficacy and engagement for all students.

The design of virtual classroom environments significantly influences gender disparities in computer science courses, impacting both course selection and anticipated success. Research by Cheryan et al. (2011) shows that altering the design of virtual classrooms from stereotypical computer science environments to more neutral or non-stereotypical settings (e.g., featuring art, nature posters) can substantially increase women's interest and perceived success in computer science. This change in environment reduces the gender gap by fostering a greater sense of belonging among female students, which is not as pronounced in male students.

Gamification can be defined as "the idea of using game design elements in non-game contexts" (Deterding et al., 2011) to further increase motivation and user activity within interaction design (Deterding et al., 2011). These game-design elements, subsequently called gamified elements, are elements often found in classical video games. However, the concept of gamification is different from designing a game, the focus lies on using the addictive component (González et al., 2014). Often used elements are points, badges, leaderboards, avatars, and narrated content. Other mechanisms include content unlocking, storytelling, and memes (Zainuddin et al., 2020). Often those elements are used in specific constellations like the PBL triad described by Werbach and Hunter (2012), which contains points, badges, and leaderboards. A system that is not only known from games, but also everyday enterprise features like loyalty programs and employee competitions (Werbach & Hunter, 2012).

- Points, because they add an absolute scale, allowing for quantifiable measurement of user achievements (Hamari et al., 2014).
- Badges, because they represent a status symbol and work like a temporary goal to strive toward, often reflecting mastery or achievement (González et al., 2014).
- Leaderboards, to compare oneself to peers, which can motivate through social

comparison but may also demotivate if not designed carefully (Almeida et al., 2021; Hamari et al., 2014).

- Avatars, as they allow users to customize their virtual representation, enhancing their identification with the activity and increasing engagement (González et al., 2014).
- Narrated content, which uses storytelling to provide context to activities, thus enriching the user's experience by embedding tasks within an appealing story (González et al., 2014).

Narrated content or storytelling and avatars are particularly interesting as there appears to be comparatively little research available on these elements, unlike the more extensively studied points, badges, and leaderboards, which was observed during the literature review. One of the positive effects of gamification is brought by the feedback in different forms (task, process, self-regulation, self) either immediate or delayed. Feedback is one of the most important factors in the relation between education and learning (Sailer & Homner, 2020). The use of gamified elements showed positive outcomes in multiple studies, in general (Hamari et al., 2014) as well as in education specific contexts (Sailer & Homner, 2020). Intelligent tutoring systems are often enhanced with gamified elements (González et al., 2014) and have shown to improve motivation and enjoyment (Jackson & McNamara, 2013). Implementing gamified elements into tutoring systems, while showing positive impact overall, could also lead to negative outcome, for example demotivation while being part of a leaderboard (Almeida et al., 2021). Almeida et al. (2021) also found in their systematic mapping, 77 papers mentioned negative effects like cheating, lack of understanding and most often lack of effect. Moreover, "Pavlovication" as Klabbers (2018) calls Gamification, is often a short question-answer-reward-cycle, conditions the user to learn conditional and narrows the possible ways to solve a problem down (Klabbers, 2018). Some studies also suggested that gamified learning platforms also lack individualism regarding choice and display of gamification elements, resulting in discomfort and negative emotions (Santos et al., 2023). To combat this missing individualism, Dehghanzadeh et al. (2024) and Oliveira et al. (2023) suggest using more independent variables to tailor the use of gamification elements. Multiple meta analyses came to the conclusion, further investigation on the effects of different gamification elements on individuals is needed Dehghanzadeh et al. (2024), Hamari et al. (2014), and Oliveira et al. (2023), as future systems could also use more individualized data to further enhance the experience of the gamified elements inside the tutoring system. Factors mentioned by Dehghanzadeh et al. (2024) include gender and age, Oliveira et al. (2023) mentions culture and gender. All studies also highlight the importance of the context, in which the learner is exposed to gamified elements.

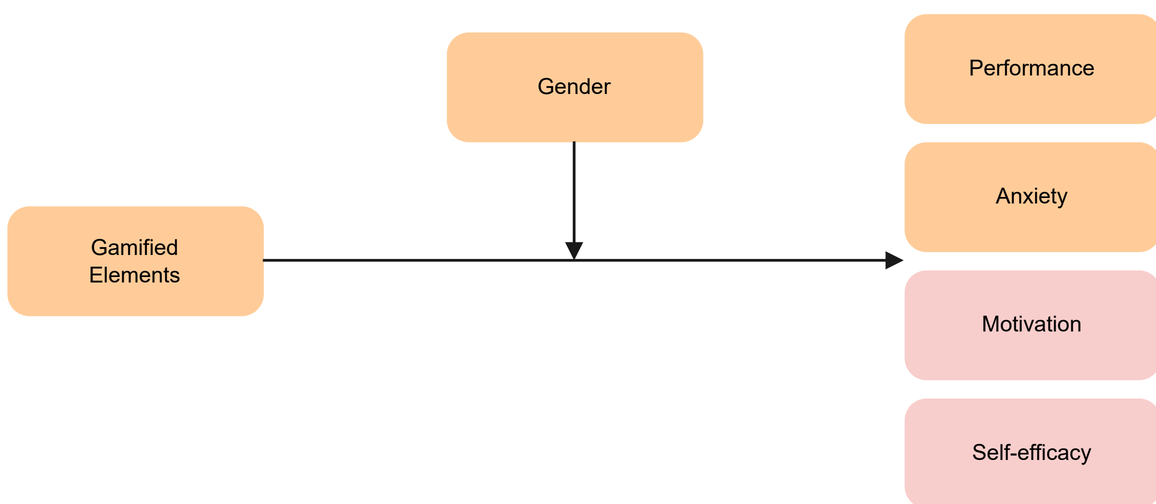
This thesis aims to explore the effects of different gamified elements and their combinations on performance and anxiety in a digital learning environment, especially focusing on the influence of gender on these effects. We believe, Gender is of particular interest, as it is a factor that is often known beforehand and already has shown to be a great influence in learning.

As such, our study is based on the work of Albuquerque et al. (2017). It explored the impact of gender stereotype threats in gamified educational environments. Their research demonstrated that male-dominated gamified contexts significantly increased anxiety levels among participants. This was particularly evident among female participants, who reported increased anxiety in such environments, potentially affecting their academic performance and engagement. The study employed a methodologically robust design involving a pretest-posttest setup and a gamified logic quiz to simulate a learning environment. These findings underline the critical influence of gender considerations in the design of educational technologies, emphasizing the need to create inclusive environments that do not inadvertently perpetuate gender biases.

This thesis will assess whether different gamified Elements can enhance the learning experience without disadvantaging any gender group. The primary goal is to design digital tutoring systems that fully leverage the potential of gamification to benefit all users equally, fostering an inclusive and effective educational environment.

Hypotheses

As noted in the first chapter, there are open questions regarding the efficiency of various gamified elements and how different genders relate to these gamified elements. The question of the efficiency of certain elements and combinations of elements remains unresolved (Dehghanzadeh et al., 2024). To explore the connection between gender and gamification elements, we have created the following model:



This model additionally incorporates concepts of motivation and self-efficacy, which, although not featured in my thesis, are included in the doctoral thesis of **Nadine Koch**.

Since males perform better than females in solving progressive matrices from age 15 onward, Hypothesis **H1a** one-sidedly formulated (**ravenStandardProgressiveMatrices2003**). The hypotheses we want to investigate in this work are:

H1 Males and females differ in their cognitive and affective states.

a) Male performance is better compared to female.

b) Male and female students differ regarding their anxiety levels.

c) Male and female students differ regarding their motivation.

d) Males have a higher self-efficacy compared to females.

H2 Different gamified elements have a varying impact on the cognitive and affective states.

a) Gamified elements impact performance differently.

b) Different gamified elements impact anxiety levels differently.

c) Different gamified elements impact motivation differently.

d) Different gamified elements impact self-efficacy differently.

H3 Different gamified elements differently impact the cognitive and affective states of males and females.

a) The influence of different gamified elements on performance differs between males and females.

b) The influence of different gamified elements on anxiety levels differs between males and females.

c) The influence of different gamified elements on motivation differs between males and females.

d) The influence of different gamification elements on self-efficacy differs between males and females.

All hypothesized effects result from interacting with the gamified digital learning environment.

Methods

Participants

119 participants were recruited from the university campus, with 69.74% identifying as ’male’, 28.57% as ’female’ and 1.68% as ’other’. Participants were aged between 18 and 27, the average age was 22 years, with a standard deviation of 2.4 years. The study programs most frequently represented were computer science, software engineering and aerospace engineering. Participants each received €15 as monetary compensation for their involvement in the study.

Design

This study explored the impact of various gamified elements and participant gender on performance and anxiety. The study design was two factorial with the factors gamified elements and gender. The first factor and independent variable was gamified elements, with participants randomly assigned to one of eight conditions: Avatars (A), Badges (B), Points (P), Leaderboards (L), Narrated Content (N), combinations of Points, Badges, Leaderboards, and Avatars (PBLA), Points, Badges, Leaderboards, Avatars, and Narrated Content (PBLAN), and no gamified elements. The other factor gender was measured in male, female and other. Each participant experienced three distinct conditions, which were sent by the server out of a randomized pregenerated batch, ensuring that all conditions were evenly distributed across participants. Participants underwent a series of tests in a fixed order during each round, beginning with a gamified performance test in a digital learning environment followed by non-gamified assessments for anxiety, self-efficacy, and motivation. At the end participants were given a monetary compensation of €15. The performance tests utilized standard progressive matrices, adapted with gamification techniques to engage and challenge participants uniquely in each round. The dependent variables included:

- **Performance**, assessed through accuracy and response times in the gamified progressive matrices.
- **Anxiety**, evaluated using a standardized questionnaire immediately after the performance test. Anxiety was measured using a shortened form of the State-Trait Anxiety Inventory (STAI) with 6 items (Marteau & Bekker, 1992).

This research employed a repeated-measures design, where each participant was exposed to three different gamification conditions chosen randomly. This within-subjects approach facilitated the analysis of individual responses to each condition across the different rounds, providing insights into how variations in gamification can affect psychological states and performance.

Materials

Physical environment

The study was conducted in two separate rooms in the cellar of a university building, one equipped with five and one with seven iMacs. As Christy and Fox (2014) suggested that the physical environment can influence the results, both rooms were equipped with the same furniture and lighting and were furnished very dry, like a typical software laboratory.

Virtual environment

The software used in this study was built by the author using SvelteKit in frontend and KTor in backend. Its UI was designed after the study by Albuquerque et al. (2017). On the iMacs the study was displayed full-screen mode using the Safari web browser to ensure no further distractions. The study consisted of 4 screens. A consent screen gave an overview and explained the data collection to the user. The consent screen had to be accepted in order to proceed. A personal detail screen collected said data; gender, age and study program. Participants also had to enter a deletion code in order to request their data's deletion after the collection.

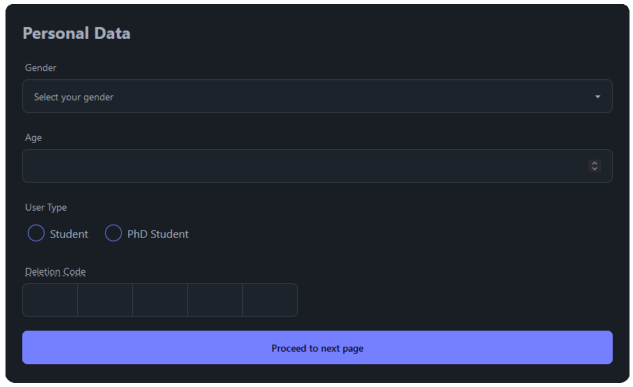


Figure 1

The personal details collection form

The next screen was the gamified learning environment, where the participants had to solve 20 questions in a row while being exposed to the gamified elements. The matrices were taken from Albuquerque et al. (2017), to generate 60 questions out of the 20, the 40 questions for iteration two and three were slightly altered versions of the original 20 made by this author.

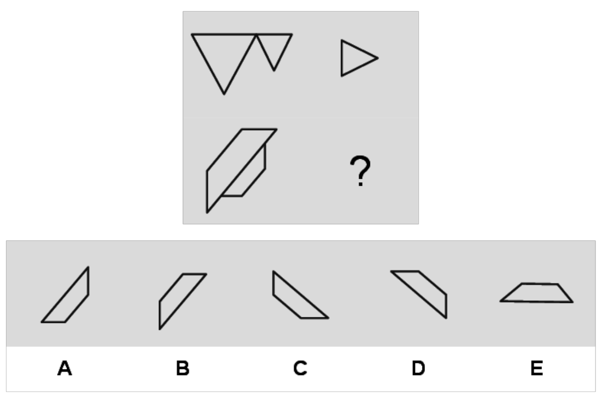
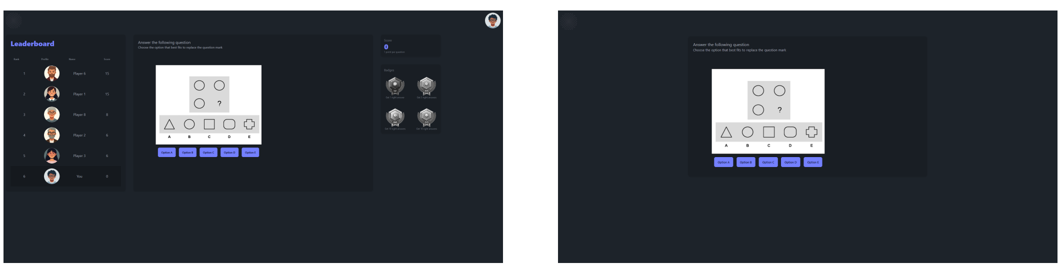


Figure 2

A standard progressive matrix, one of the tasks given to the participants

The gamified learning environment consisted of different UI elements representing the gamified elements.

- **Leaderboards**: A list of participants and scores, including the current participant. The other players shown were not real.
- **Badges**: An array of four badges that were awarded for 1, 5, 10 and 18 correctly answered questions.
- **Avatars**: A small avatar that was shown in the top right corner of the screen and on the leaderboard. To increase identification with the avatar further, the participants were asked to choose one of 15 different avatars before the iteration.
- **Narrated content**: Narrated content was shown in the bottom right corner of the screen. It was presented as a speech bubble with an avatar next to it, in case avatars were enabled. It showed a random praise or encouragement sentence every three questions.
- **Points**: A counter next to the question frame showed the current points. One point was awarded for each correctly answered question.



(a) *The Digital Learning Environment with Points, Badges, Leaderboards and Avatars enabled. Narrated content in only shown between questions.*

(b) *The Digital Learning Environment with no Gamified Elements enabled*

Figure 3

Comparison of the Digital Learning Environment with and without gamified elements enabled

After the gamified learning environment, the participants were shown a questionnaire for anxiety, motivation and self-efficacy. The three questionnaires were a six-question shortened form of the State-Trait Anxiety Inventory (STAI) (Marteau & Bekker, 1992), the eight-question General Self-Efficacy Scale (GSE) (Guay et al., 2000) and the 16-question Situational Intrinsic Motivation Scale (SIMS) (Chen et al., 2001).

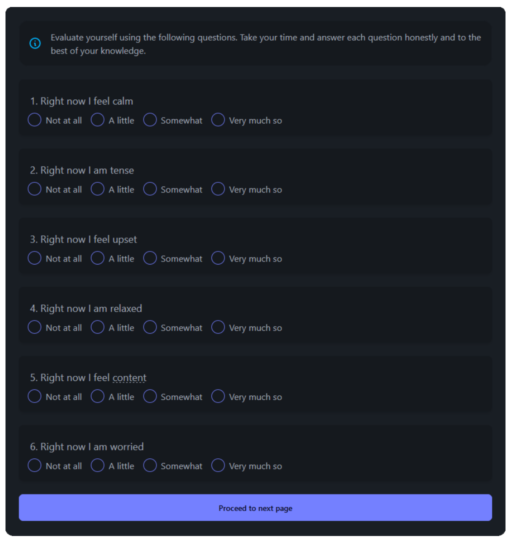
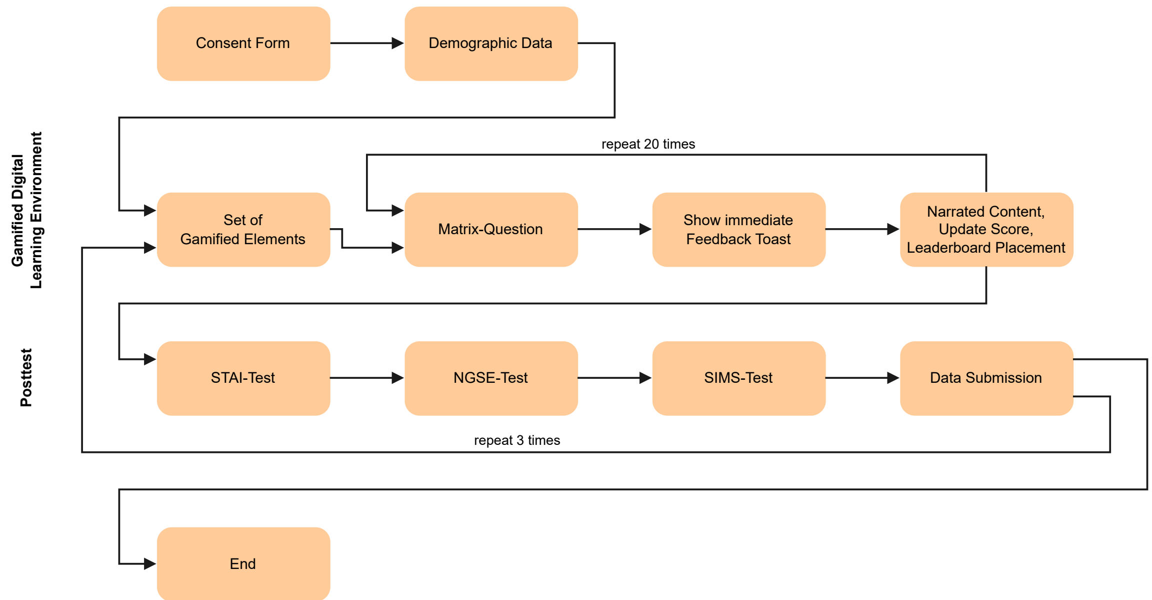


Figure 4

The anxiety questionnaire

To submit data to the backend there was a data submission screen that guided the participants to the next iteration or in case of the third iteration to the end of the study.

Procedure



Participants have been enlisted at the university campus and invited to engage in a study concerning gamification, with an incentive of 15€ offered upfront for their involvement. Following a brief overview of the study's framework, they have been directed to select both a room and a computer. The initial screens presented were those seeking consent and outlining the study details. Subsequently, participants have inputted their data, leading into a series of three questioning phases. Each phase has initiated with the gamified learning environment, followed by three questionnaires, and concluded with a data submission interface. After answering one question the next question has a one-second delay which increases to four seconds if narrated content is shown. The sequence and consistency of the testing procedure, including the series of questions asked in the gamified digital learning environment have always been maintained to ensure the reliability of measurements and comparability of results across the various stages of the experiment. Participants have been advised to proceed at their own pace and refrain from communicating with fellow participants throughout the duration of the study. Upon completion of the third iteration, they have been acknowledged for their contribution and compensated with the 15€.

Scoring

The data was cleaned and processed before analysis to ensure accuracy and reliability. The scores for the different conditions were calculated as follows:

- **Performance:** Performance was calculated as the ratio of correctly answered questions to the total number of questions. If this ratio was below 0.25 for any dataset, that dataset was excluded from further analysis. This threshold ensures that only participants with a sufficient level of engagement are included in the study.
- **State-Trait Anxiety Inventory (STAI):** The STAI scores were calculated using the six-item short-form version of the State scale, as developed by Marteau and Bekker (1992). Participants rated their responses on a scale from "Not at all" to "Very much so," corresponding to numerical values from zero to five. The original test contains both anxiety-present and anxiety-absent items, with appropriate weights applied as per the guidelines by Marteau and Bekker (1992). Negative weights were used for items indicating anxiety (e.g., "Right now I am worried").
- **New General Self-Efficacy Scale (NGSE):** The NGSE scores were calculated based on the method outlined by Chen et al. (2001). Participants' responses were averaged to obtain the mean score, which indicates their overall self-efficacy across different situations. This scale, validated by Chen et al. (2001), captures the general sense of self-efficacy.
- **Situational Motivation Scale (SIMS):** The SIMS was used to measure participants' situational motivation in the gamified test environment. This scale assesses intrinsic motivation, identified regulation, external regulation, and amotivation, providing a comprehensive view of the participants' motivational states. Scores were calculated by averaging the participants' responses, as suggested by Guay et al. (2000).

Evaluation

Data Exclusion

Participants who identified their gender as "other" were excluded from the analysis because the present research available and used for this thesis primarily focuses on comparisons between male and female participants. This criterion led to the exclusion of two participants. Additionally, participants who achieved less than 25% correct answers in the gamified learning environment were excluded. This threshold was set because there were five possible answers for each question, and random clicking would statistically result in a 20% correct response rate. Therefore, a performance below 25% suggests either random guessing or a fundamental misunderstanding of the task. Furthermore, any incomplete data sets were excluded to ensure the integrity and consistency of the analysis, resulting in the exclusion of one additional participant. Initially, there were 120 data sets, and after applying the exclusion criteria, 117 data sets remained.

Outline of statistical analysis

Having preprocessed and cleaned the data, we proceeded with our statistical analysis using linear mixed models (LMMs). These models were chosen for their ability to handle the complexities of repeated measures from the same subjects under varying conditions. By incorporating both fixed effects (gender and gamified elements) and random effects (individual differences), LMMs provided a robust framework for our analysis.

We utilized the Nelder-Mead optimization method to estimate the parameters of our models. This method is ideal for our needs as it efficiently handles models with multiple interacting effects without requiring derivative calculations, making it suitable for our complex dataset.

For the estimation of variance components within our models, we employed

Restricted Maximum Likelihood (REML). REML is preferred in mixed model contexts because it adjusts the estimates for the fixed effects, providing unbiased variance estimates despite the presence of random effects.

Finally, to ensure accurate inference regarding the fixed effects, we applied the Satterthwaite approximation for estimating degrees of freedom. This method helps in achieving more reliable p-values by adjusting the degrees of freedom for the complexity of the model, crucial in cases with multiple levels of interactions and a limited sample size.

This combination of methods and their implementation through LMMs allowed us to systematically analyse the effects of gender and gamified elements on performance and anxiety, controlling for individual variability and the specifics of the experimental design.

Results

Performance

Our analysis revealed no significant main effects or interactions for all hypotheses regarding performance, as documented in Table **??**. Despite this lack of significant effects, some descriptive effects are notable.

Women exhibited lower performance levels when leaderboards were the gamified Element used (M= .700, SE = .056) compared to men (M = .835, SE = .024) and to the overall average performance in gamified settings for women (M = .846, SE = 0.030). Notable is also the variability suggested by the large standard error for women in this leaderboard condition.

Anxiety

Similarly, our analysis revealed no significant main effects or interactions for all hypotheses regarding anxiety levels, as detailed in Table **??**. Nonetheless, some

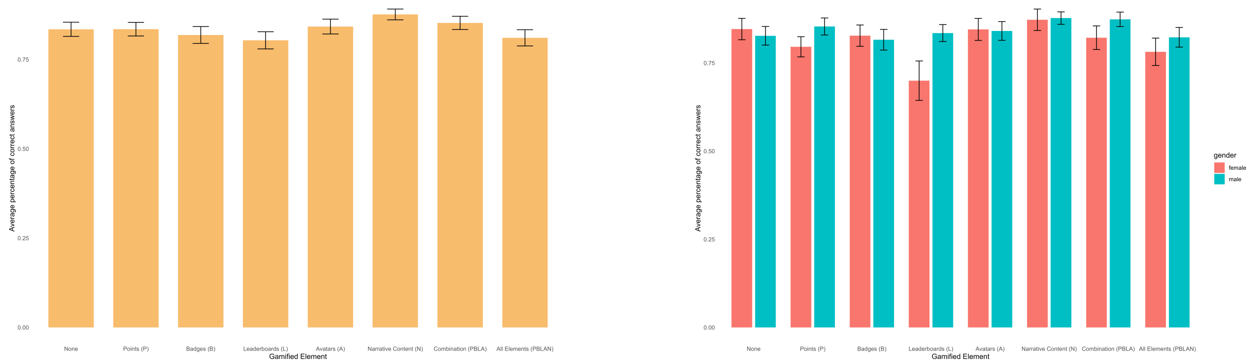


Figure 5

On the left: Overall performance across different gamification elements. On the right: Performance differences between genders across different gamified elements.

descriptive trends are worth noting.

Figure **??** shows that women experienced higher anxiety levels in the non-gamified environment (M = .571, SE = .092) and with leaderboards (M = .683, SE = .205) compared to men in the same settings (M = .296, SE = .099 and M = .566, SE = .119, respectively). Conversely, the use of avatars led to lower anxiety levels in women (M = .187, SE = .095) compared to men (M = .476, SE = .102).

Discussion

Summary of Research focus

This thesis explored the effects of gamification elements on performance and anxiety in digital learning environments, with a focus on gender differences. Using a randomized controlled design, the study assessed various gamification components like points, badges, leaderboards, and avatars, analysing responses from male and female participants. The main goal was to identify how different genders react to these gamification strategies regarding their cognitive and affective states.

Although the study did not find statistically significant effects, the data trends were

Table 1

Results of the linear mixed model analysis for percentage correct effects.

Variable	std. Beta β	standardized 95% CI of β	p	Statistic	df
Gender [male]	0.07	[-0.47, 0.62]	0.994	0.27	277.62
Gamified Element [Points (P)]	-0.33	[-0.82, 0.16]	0.493	-1.35	211.82
Gamified Element [Badges (B)]	0.06	[-0.50, 0.62]	0.994	0.21	214.27
Gamified Element [Level (L)]	-0.48	[-1.04, 0.09]	0.384	-1.67	213.69
Gamified Element [Avatars (A)]	0.13	[-0.39, 0.64]	0.994	0.49	208.19
Gamified Element [Narrative Content (N)]	0.44	[-0.07, 0.95]	0.384	1.70	207.03
Gamified Element [Combination (PBLA)]	0.01	[-0.45, 0.48]	0.994	0.06	208.37
Gamified Element [All Elements (PBLAN)]	0.06	[-0.46, 0.57]	0.994	0.22	210.22
Gender [male] × Gamified Element [Points (P)]	0.42	[-0.20, 1.04]	0.493	1.33	214.33
Gender [male] × Gamified Element [Badges (B)]	-0.22	[-0.88, 0.45]	0.994	-0.64	213.78
Gender [male] × Gamified Element [Level (L)]	0.61	[-0.06, 1.28]	0.384	1.79	214.82
Gender [male] × Gamified Element [Avatars (A)]	0.01	[-0.62, 0.64]	0.994	0.03	209.52
Gender [male] × Gamified Element [Narrative Content (N)]	-0.00	[-0.63, 0.63]	0.994	-0.01	210.19
Gender [male] × Gamified Element [Combination (PBLA)]	0.37	[-0.25, 0.98]	0.548	1.18	211.59
Gender [male] × Gamified Element [All Elements (PBLAN)]	0.01	[-0.63, 0.64]	0.994	0.02	210.43

in line with existing research, indicating consistent patterns in how participants reacted to the gamification elements. This suggests that while the effects were not strong enough to be statistically significant, the observed behaviours reflected known theories and previous studies. These results highlight the challenges of isolating the impacts of gamification and gender in educational contexts, pointing to the potential need for further research with larger sample sizes or different approaches.

Interpretation of Results

The absence of statistically significant findings in this study does not negate the observation of certain trends that suggest a subtle influence of gamification on learning outcomes. These observed trends align with previous research that also noted similar patterns without definitive effects (Dehghanzadeh et al., 2024; Hamari et al., 2014). This consistency suggests that while the direct impact of gamified elements may not be significant in our study, there is a consistent directional influence on learners' cognitive

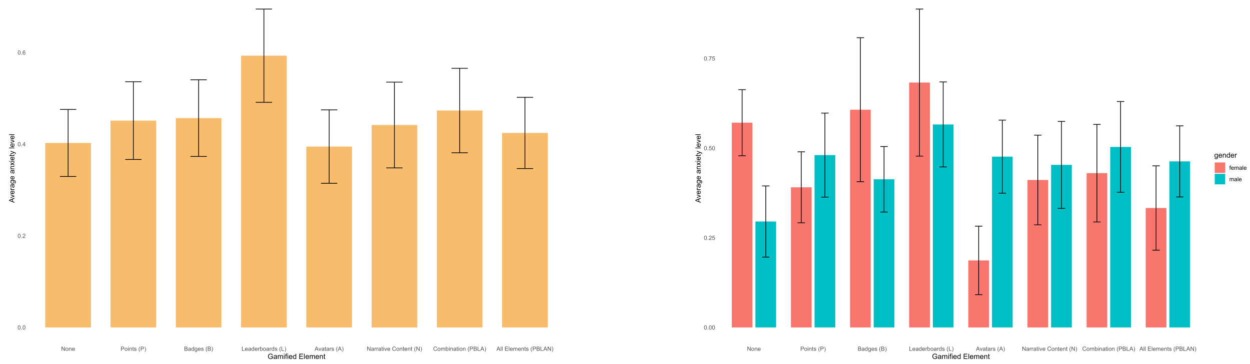


Figure 6

On the left: Anxiety levels across different gamification elements. On the right: Differences in anxiety levels between genders in various gamification contexts.

and affective states. This highlights the nuanced nature of gamification effects, suggesting that they may manifest under specific conditions or within certain contexts, as Dehghanzadeh et al. (2024), Koivisto and Hamari (2019), and Oliveira et al. (2023).

Implications for Theory and Practice

Theoretically, the observed trends, hint at potential variations in how gamification impacts different learners, suggesting a need for broader conceptual models. This could inspire further research into personalized learning environments. Practically, these preliminary findings encourage a cautious approach to integrating gamification in educational settings. Designers might consider more flexible gamification systems that can be adjusted based on learner feedback, potentially enhancing user engagement without a one-size-fits-all strategy.

Limitations

This study has several key limitations that must be acknowledged. Firstly, the reliance on self-reported data from students may lead to potential biases and inaccuracies in the findings. Previous research suggests that objective measures, such

Table 2

Results of the linear mixed model analysis for STAI effects.

Variable	std. Beta β	standardized 95% CI of β	p	Statistic	df
Gender [male]	-0.25	[-0.83, 0.34]	0.623	-0.83	288.25
Gamified Element [Points (P)]	-0.28	[-0.83, 0.27]	0.564	-1.00	216.76
Gamified Element [Badges (B)]	0.23	[-0.40, 0.86]	0.623	0.73	221.07
Gamified Element [Leaderboards (L)]	-0.04	[-0.67, 0.59]	0.913	-0.11	220.38
Gamified Element [Avatars (A)]	-0.62	[-1.20, -0.04]	0.228	-2.11	213.37
Gamified Element [Narrative Content (N)]	-0.53	[-1.10, 0.05]	0.241	-1.81	211.70
Gamified Element [Combination (PBLA)]	-0.03	[-0.55, 0.50]	0.913	-0.11	213.53
Gamified Element [All Elements (PBLAN)]	-0.60	[-1.18, -0.02]	0.228	-2.04	215.77
Gender [male] × Gamified Element [Points (P)]	0.42	[-0.28, 1.11]	0.481	1.18	220.20
Gender [male] × Gamified Element [Badges (B)]	-0.18	[-0.93, 0.57]	0.787	-0.47	220.26
Gender [male] × Gamified Element [Leaderboards (L)]	0.28	[-0.47, 1.04]	0.623	0.74	221.54
Gender [male] × Gamified Element [Avatars (A)]	0.64	[-0.07, 1.35]	0.241	1.79	214.84
Gender [male] × Gamified Element [Narrative Content (N)]	0.50	[-0.21, 1.21]	0.437	1.40	215.47
Gender [male] × Gamified Element [Combination (PBLA)]	0.10	[-0.58, 0.79]	0.882	0.29	217.43
Gender [male] × Gamified Element [All Elements (PBLAN)]	0.45	[-0.26, 1.17]	0.481	1.25	215.91

as sensor-based tracking, could provide more reliable data (Woolf et al., 2010). Another significant limitation arises from the demographic and location constraints of our sample. The majority of participants were recruited from the computer science building on a technical campus, which may not provide a diverse representation of the general student population. Additionally, the customization options for avatars were limited, lacking diverse representations such as Hijabs or beards, as noted by participants. This limitation may affect the engagement and identification of users with the avatars, potentially skewing the results regarding the impact of gamification on different genders. The narrated content of the gamified elements could also be not engaging enough, being confined to a small portion of the screen without integrating characteristics, abilities, or dialogues that could enhance user interaction. This might have contributed to some participants opting to disable gamified elements, thus not experiencing the intended gamified environment fully. Furthermore, the design of the study involved three iterations per participant, which might have led to habituation effects. These effects could influence the outcomes in terms of learning environment adaptability, anxiety, motivation, and self-efficacy, thus potentially diminishing the study's ability to measure these constructs accurately over time. Finally, the lack of sufficient individualized feedback post-study was highlighted as a concern. Future iterations of this study should consider incorporating mechanisms for more personalized and actionable feedback to enhance the learning and adaptation process for participants. Despite these limitations, the findings provide valuable insights into the intersection of gamification and gender, suggesting avenues for future research and practical applications in educational environments.

Future Research

As highlighted in the introduction, the exploration of individual characteristics in the selection of gamified elements within learning environments requires further investigation. This thesis has laid foundational insights into how gender influences engagement with gamified systems. However, other dimensions such as age, culture, and notably, the educational level (ranging from school-level to higher education and adult learning) merit in-depth exploration to comprehensively understand the dynamics at play.

Moreover, while the study's learning environment provided a basic platform, it does not directly represent the probable application of gamified elements. As noted earlier, gamified elements are prevalently integrated into Intelligent Tutoring Systems. Therefore, conducting studies that focus on gamified ITS can yield more pertinent insights into their efficacy and application. As the modelling of such systems, even without gamification, is quite extensive and customizable per learner itself, for it to work properly, it should be tested in an ITS.

Additionally, the temporal scope of this study was limited, addressing only short-term interactions. Prior research has established the significance of longitudinal studies in this domain (Dehghanzadeh et al., 2024; Oliveira et al., 2023). Extending research timelines to span multiple semesters would provide a more robust understanding of the long-term impacts of gamification on learning outcomes and student engagement. Such extended studies are quite challenging in every research area, but crucial for observing changes in learner stategies and behaviour and the sustainability of gamification benefits over time.

In conclusion, advancing research in these areas will not only enrich our understanding of how different groups respond to gamified learning but also enhance the design and implementation of educational technologies that are inclusive and effective across diverse learning environments.

Conclusion

This study's exploration into the effects of gamification elements on performance and anxiety with a focus on gender differences has provided initial insights that, while not statistically significant, point to subtle yet potentially important trends. The observed data suggest that certain gamified elements could have differential impacts on cognitive and affective outcomes across genders. Although these effects were not strong enough to achieve statistical significance in this study, the descriptive findings hint at underlying patterns that merit further investigation.

The consistency of these patterns with prior research indicates that with improved research designs that address the limitations of this study, significant findings could be obtained. These findings could be very important for advancing the understanding of how gamification can be optimized to enhance performance and anxiety effectively. Future research should consider these aspects to uncover more robust evidence that can contribute to the ongoing discourse in educational technology.

Acknowledgements

We acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation under the DFG reference number **UP 31/1**) for the

Stuttgart Research Focus *Interchange Forum for Reflecting on Intelligent Systems (IRIS)*.

References

Albuquerque, J., Bittencourt, I. I., Coelho, J. A., & Silva, A. P. (2017). Does gender stereotype threat in gamified educational environments cause anxiety? An experimental study. *Computers & Education*, *115*, 161–170. https://doi.org/10.1016/j.compedu.2017.08.005

Almeida, C., Kalinowski, M., & Feijó, B. (2021). A Systematic Mapping of Negative Effects of Gamification in Education/Learning Systems. *2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, 17–24. https://doi.org/10.1109/SEAA53835.2021.00011

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a New General Self-Efficacy Scale. *Organizational Research Methods*, *4*(1), 62–83. https://doi.org/10.1177/109442810141004

Cheryan, S., Meltzoff, A. N., & Kim, S. (2011). Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes. *Computers & Education*, *57*(2), 1825–1835. https://doi.org/10.1016/j.compedu.2011.02.004

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

Dabaj, F. (2009, April). *The Role of Gender and Age on Students' Perceptions towards Online Education Case Study: Sakarya University, Vocational High School* (No. 2). Retrieved July 22, 2024, from https://eric.ed.gov/?id=ED505933 ERIC Number: ED505933.

De Santana, S. J., Paiva, R., Bittencourt, I. I., Ospina, P. E., De Amorim Silva, R., & Isotani, S. (2016). Evaluating the Impact of Mars and Venus Effect on the Use of an Adaptive Learning Technology for Portuguese and Mathematics. *2016 IEEE*

16th International Conference on Advanced Learning Technologies (ICALT), 31–35. https://doi.org/10.1109/ICALT.2016.58

Dehghanzadeh, H., Farrokhnia, M., Dehghanzadeh, H., Taghipour, K., & Noroozi, O. (2024). Using gamification to support learning in K-12 education: A systematic literature review. *British Journal of Educational Technology*, *55*(1), 34–70. https://doi.org/10.1111/bjet.13335

Dermeval, D., Albuquerque, J., Bittencourt, I. I., Isotani, S., Silva, A. P., & Vassileva, J. (2019). GaTO: An Ontological Model to Apply Gamification in Intelligent Tutoring Systems. *Frontiers in Artificial Intelligence*, *2*. https://doi.org/10.3389/frai.2019.00013

des statistischen Bundesamts, P. (2020, December 8). *Digitales Lernen nimmt stark zu*. Statistisches Bundesamt. Retrieved July 20, 2024, from https://www.destatis.de/DE/Presse/Pressemitteilungen/2020/12/PD20_N081_63.html

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification".

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

El Hadbi, A., Hachem, E. K., Bourray, H., Rziki, M. H., & Oubalahcen, H. (2024). Review Study on the Adaptive and Personalized Intelligent Tutoring Systems in E-learning. In M. Ezziyyani, J. Kacprzyk, & V. E. Balas (Eds.), *International Conference on Advanced Intelligent Systems for Sustainable Development (AI2SD'2023)* (pp. 45–68). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54288-6_5

García-Morales, V. J., Garrido-Moreno, A., & Martín-Rojas, R. (2021). The Transformation of Higher Education After the COVID Disruption: Emerging

Challenges in an Online Learning Scenario. *Frontiers in Psychology*, *12*, 616059. https://doi.org/10.3389/fpsyg.2021.616059

González, C., Mora, A., & Toledo, P. (2014). Gamification in intelligent tutoring systems. *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality*, 221–225. https://doi.org/10.1145/2669711.2669903

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, *24*(3), 175–213. https://doi.org/10.1023/A:1005614228250

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? — A Literature Review of Empirical Studies on Gamification. https://doi.org/10.1109/HICSS.2014.377

Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology*, *105*(4), 1036–1049. https://doi.org/10.1037/a0032580

Kampshoff, M., & Wiepcke, C. (Eds.). (2012). *Handbuch Geschlechterforschung und Fachdidaktik*. VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-18984-0

Klabbers, J. H. (2018). On the Architecture of Game Science. *Simulation & Gaming*, *49*(3), 207–245. https://doi.org/10.1177/1046878118762534

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

Lindqvist, A., Sendén, M. G., & Renström, E. A. (2021). What is gender, anyway: A review of the options for operationalising gender. *Psychology & Sexuality*.

Retrieved May 18, 2024, from https://www.tandfonline.com/doi/abs/10.1080/19419899.2020.1729844

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, *31*(3), 301–306. https://doi.org/10.1111/j.2044-8260.1992.tb00997.x

Nkambou, R., Bourdeau, J., & Mizoguchi, R. (Eds.). (2010). *Advances in Intelligent Tutoring Systems* (Vol. 308). Springer. https://doi.org/10.1007/978-3-642-14363-2

Oliveira, W., Hamari, J., Shi, L., Toda, A. M., Rodrigues, L., Palomino, P. T., & Isotani, S. (2023). Tailored gamification in education: A literature review and future agenda. *Education and Information Technologies*, *28*(1), 373–406. https://doi.org/10.1007/s10639-022-11122-4

Psotka, J., Massey, L. D., & Mutter, S. A. (**typeredactors**). (1988). *Intelligent tutoring systems: Lessons learned*. Lawrence Erlbaum Associates, Inc.

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

Santos, J., Andrade, E., Benevides, K., Silva, K., Nascimento, J., Bittencourt, I., Pereira, M., Fernandes, S., & Isotani, S. (2023). Does gender stereotype threat affects the levels of aggressiveness, learning and flow in gamified learning environments?: An experimental study. *Education and Information Technologies*, *28*(2), 1637–1662. https://doi.org/10.1007/s10639-022-11220-3

Swacha, J. (2021). State of Research on Gamification in Education: A Bibliometric Survey. *Education Sciences*, *11*(2), 69. https://doi.org/10.3390/educsci11020069

Wehrwein, E. A., Lujan, H. L., & DiCarlo, S. E. (2007). Gender differences in learning style preferences among undergraduate physiology students. *Advances in Physiology Education*, *31*(2), 153–157. https://doi.org/10.1152/advan.00060.2006

Werbach, K., & Hunter, D. (2012, January 1). *For the Win: How Game Thinking can Revolutionize your Business*.

Woolf, B. P., Arroyo, I., Cooper, D., Burleson, W., & Muldner, K. (2010). Affective Tutors: Automatic Detection of and Response to Student Emotion. In R. Nkambou, J. Bourdeau, & R. Mizoguchi (Eds.), *Advances in Intelligent Tutoring Systems* (pp. 207–227). Springer. https://doi.org/10.1007/978-3-642-14363-2_10

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

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0