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:	06. Mai 2024
Beendet am:	06. September 2024

Abstract

Diese Bachelorarbeit untersucht die Auswirkungen verschiedener Gamification-Elemente auf Leistung und Angst in digitalen Lernumgebungen, mit einem besonderen Fokus auf Geschlechterunterschiede. Trotz der breiten Anwendung von Gamification in Bildungstechnologien ist das Verständnis darüber, wie individuelle Faktoren wie Geschlecht diese Effekte beeinflussen, begrenzt. Die Studie verwendet ein 2×8 faktorielles Design, um die Interaktionen zwischen Gamification-Elementen und Geschlecht zu analysieren, wobei 117 Teilnehmer aus verschiedenen Studiengängen rekrutiert wurden. Die Ergebnisse zeigen, dass verschiedene Gamification-Elemente die Leistung und Angst unterschiedlich beeinflussen, jedoch waren diese Gruppenunterschiede überwiegend deskriptiv und nicht signifikant. Diese Erkenntnisse deuten darauf hin, dass die Anpassung von Gamification-Strategien an geschlechtsspezifische Reaktionen potenziell die Effektivität von digitalen Lernumgebungen verbessern könnte. Die Studie betont die Notwendigkeit weiterer Forschungen, um die Beziehung zwischen Gamification, Geschlecht und Lernausgängen tiefer zu verstehen.

Keywords: Gamification, Geschlecht, Intelligente Tutorielle Systeme, Leistung, Angst

Abstract

This bachelor thesis explores the effects of various gamification elements on performance and anxiety in digital learning environments, with a specific focus on gender differences. Despite the widespread application of gamification in educational technologies, understanding of how individual factors such as gender influence these effects remains limited. The study employed a 2×8 factorial design to analyze interactions between gamification elements and gender, recruiting 117 participants from various degree programs. Findings indicate that different gamification elements variably affect performance and anxiety, however, these group differences were predominantly descriptive and not significant effects. These insights suggest that tailoring gamification strategies to gender-specific responses could potentially enhance the effectiveness of digital learning environments. The study underscores the need for further research to deepen understanding of the relationship between gamification, gender, and learning outcomes.

Keywords: Gamification, Gender, Intelligent tutoring systems, Performance, Anxiety

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

Digital learning environments (DLE) offer a wide range of features and flexibility to support students learning processes. Tutoring systems are one kind of digital learning environments ranging 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). DLE are in use at higher education institutions and widely researched, especially in the field of computer science (Zawacki-Richter et al., 2019).

This could be due to the often mandated transition to remote education during the COVID-19 pandemic highlighted the indispensable role of digital learning environments (García-Morales et al., 2021). Especially Tutoring Systems, although often costly and impersonalized have become an important tool for education institutions (El Hadbi et al., 2024). 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 (Statistisches Bundesamt, 2020).

Tutoring systems can be enhanced with some sort of intelligence, resulting in intelligent tutoring systems (ITS). Intelligent tutoring systems adapt dynamically to learner's needs, incorporating usage factors such as learner 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).

Jackson and McNamara (2013) found that traditional tutoring systems lead to boredom and disengagement after long periods of use. Further, they showed the use of gamified elements in tutoring systems significantly improved the motivation and performance of 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). The concept of gamification also has attracted substantial interest within the educational sciences (Swacha, 2021). These game-design elements, subsequently called gamified elements, are elements often found in video games. However, the concept of gamification is different from designing a game, the focus lies on applying the addictive component to other environments, in this case education (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). Those elements are often 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 as they provide feedback constantly (Wouters et al., 2013) showed positive outcomes in multiple studies, in general (Hamari et al., 2014) as well as in education specific contexts (Sailer & Homner, 2020).

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). Further, gamified elements are often included into DLE, as they significantly improve the learning experience (Dermeval et al., 2019), motivation and enjoyment (González et al., 2014; Jackson & McNamara, 2013). Incorporating gamified elements not only enhances the engagement and motivation within the DLE but also necessitates mechanisms for tracking progress, such as content unlocking (González et al., 2014). The evolving landscape of DLE research also includes emotional and relational dynamics, linking student emotions and teacher-student relationships to learning efficacy and motivation (Woolf et al., 2010).

However, despite the potential positive impacts, the introduction of gamified elements can also lead to negative outcomes. Almeida et al. (2021) found in their systematic mapping, 77 papers mentioned negative effects like cheating, lack of understanding, demotivation in leaderboards and most often lack of effect of the gamified elements. 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. For example, technological proficiency and communication challenges in online settings which 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).

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 in DLE as 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). Recognizing and addressing gender differences in learning 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 creating a greater sense of belonging among female students, which is not as pronounced in male students.

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). Competition, often created with leaderboards, results in higher gratification in men compared to women (Lucas & Sherry, 2004). Gamified environments also have been reported to increase tension and lower perceived competence in women compared to an perceived competence increase in men (Lais Z Pedro et al., 2015).

This can also be seen in a study by Albuquerque et al. (2017) exploring 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. Albuquerque et al. (2017) 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 Study

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. Gender could be 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). Although similar, the study did focus on stereotype threat and did not investigate the effects of different gamified elements on performance and anxiety.

Also important in this and Albuquerque et al. (2017) study's context, males perform better than females in solving progressive matrices from age 15 onward (Lynn & Irwing, 2004).

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. To explore the impact of gender and gamification elements on performance and anxiety, we have created the following model:

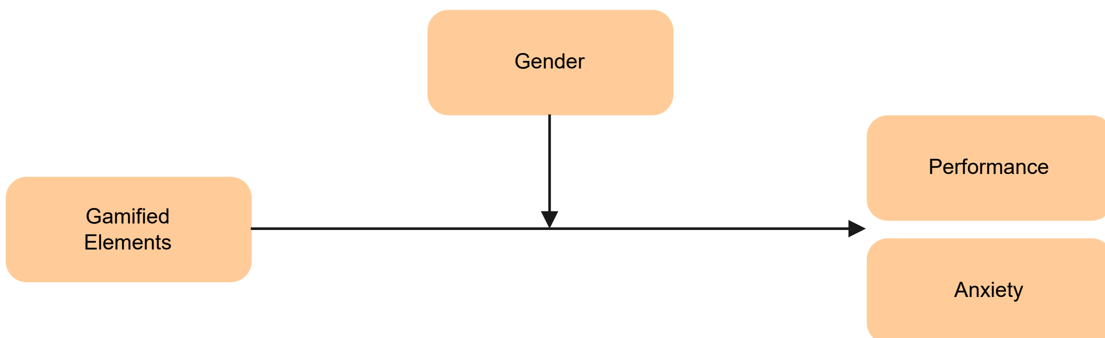


Figure 1

This diagram illustrates the impact of gamified elements (independent variable) on performance and anxiety (dependent variables), mediated by gender (mediating variable).

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.

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.

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.

Methods

Participants

117 students were recruited between the 24.06.2024 and the 15.07.2024 from the university campus, with 83 (70.94%) identifying as male, 34 (29.06%) as female. Participants were aged between 18 and 27, with a mean of 21.91 (*SD* = 2.38) years, resulting in an interval of [19.53, 24.29]. Students of 32 study programs took part with the study programs most frequently represented being technical study programs, like Computer Science B.Sc. (n = 34), Software Engineering B.Sc. (n = 11), Aerospace Engineering B.Sc. (n = 8), Mechanical Engineering B.Sc. (n = 7) and Data Science B.Sc. (n = 5). Participants gave informed consent, each received €15 as monetary compensation for their involvement in the study.

Design

This study employed a 2 $\times$ 8 factorial design, examining the impacts of gamified elements and gender on performance and anxiety [1] within a digital learning environment.

[1] The study additionally incorporated concepts of motivation and self-efficacy, which, although not featured in my thesis, are included in the doctoral thesis of Nadine Koch

The independent variables were gamified elements and gender. Gamified elements, included the following eight conditions: No Gamification (None), Points (P), Badges (B), Leaderboards (L), Avatars (A), Narrated Content (N), "Points, Badges, Leaderboards, Avatars" (PBLA) and "Points, Badges, Leaderboards, Avatars, Narrated Content" (PBLAN). The gender variable was measured categorically with participants identifying as male or female. Each participant experienced three blocks of a quiz with 20 items enhanced with one different condition (gamified element), chosen randomly from a pre-generated set to ensure a balanced distribution across conditions and three questionnaires about it afterwards. However, since each participant experienced only three out of the eight possible conditions, and different participants could experience different sets of conditions, this setup integrates elements of both within-subjects and between-subjects designs. This allowed for the analysis of interaction effects between gamified elements and gender on the dependent variables, which were performance and anxiety.

Materials

Physical Environment

The study was conducted in two separate software laboratories in the cellar of a university building, one equipped with five and one with seven iMacs. As Christy and Fox (2014) suggested that the environment can influence the results, both rooms were equipped with the same furniture and lighting and were furnished 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 7 pages. 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. The demographic data screen collected gender, age and study program. Participants also had to enter a deletion code in order to request their data's deletion after the collection. A gamified learning environment, where the participants had to solve questions while being exposed to the gamified elements. The matrices were taken from Albuquerque et al. (2017), for example Figure 2 (left), 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. Those alterations included changing the position or rotation of the elements in the matrix and the five possible answers, for example Figure 2 (right).

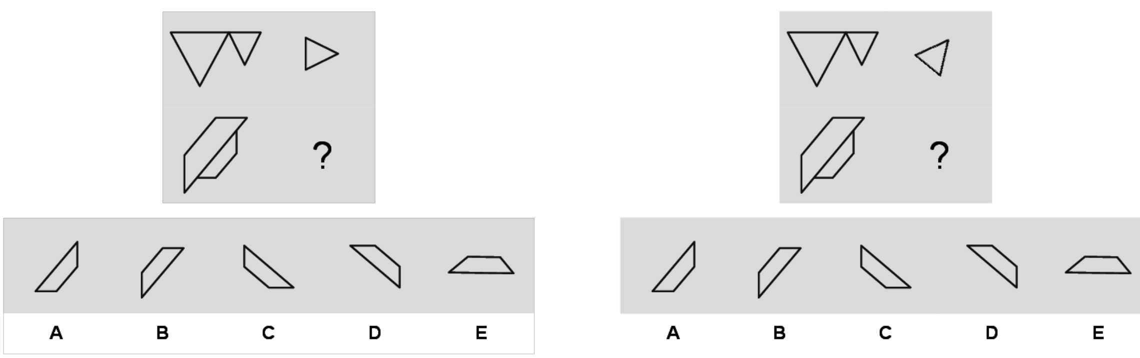


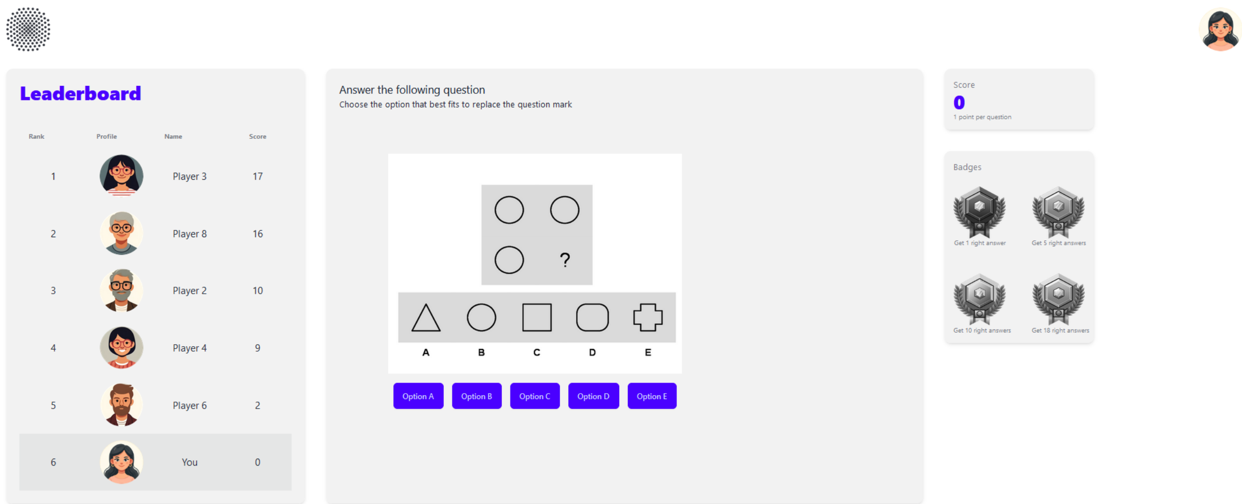
Figure 2

Left: A standard progressive matrix from Albuquerque et al. (2017) shown in the quiz to the participants with D being the right answer. Right: A slightly altered version by rotating the top right triangle of the same matrix, changing the right answer to E.

This gamified learning environment shown in Figure 3 consisted of different UI elements representing the gamified elements.

- **Leaderboards**: A list of participants and scores, including the current participant. The list items consisted of rank, name in the format "Player " + random number and score, the players item was highlighted. The other players shown were not real, their scores were randomly generated so that the player started on the sixth place and the first place was always possible.

- **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 chat bubble with an avatar next to it, appearing like it spoke, 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.



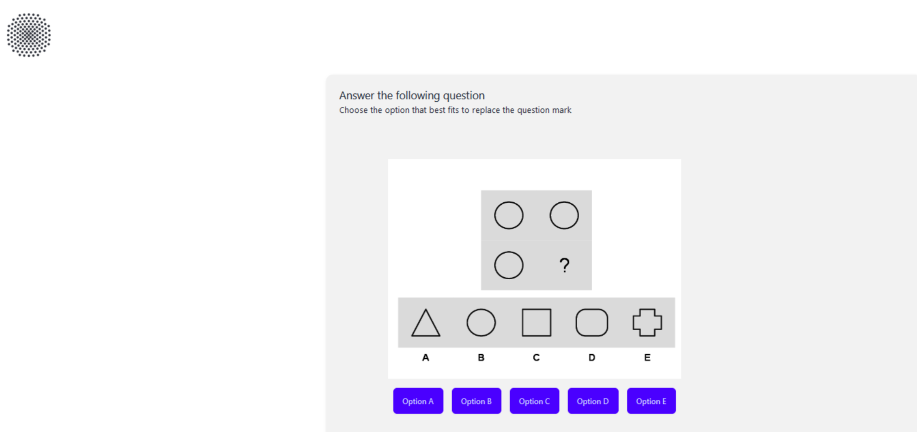


Figure 3

Comparison of the Digital Learning Environment with gamified elements enabled (left) and without gamified elements enabled (right). Narrated content in only shown between questions so it is not possible to show it in the same screenshot.

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 and 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)).

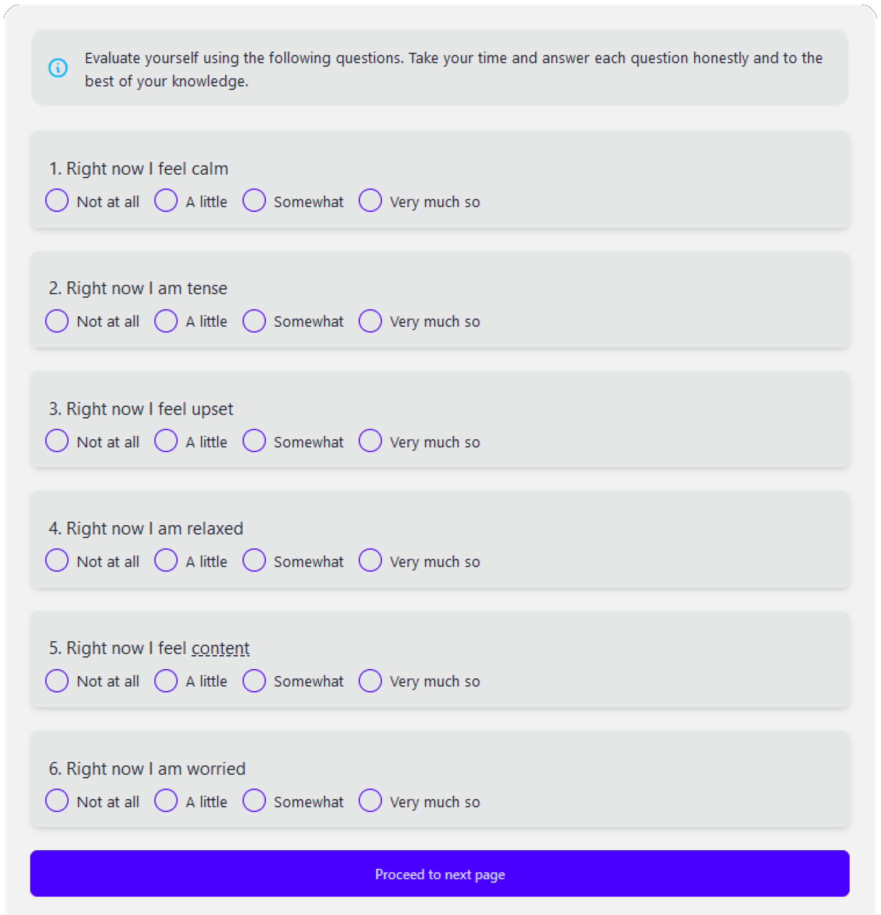


Figure 4

The anxiety questionnaire with 6 items.

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

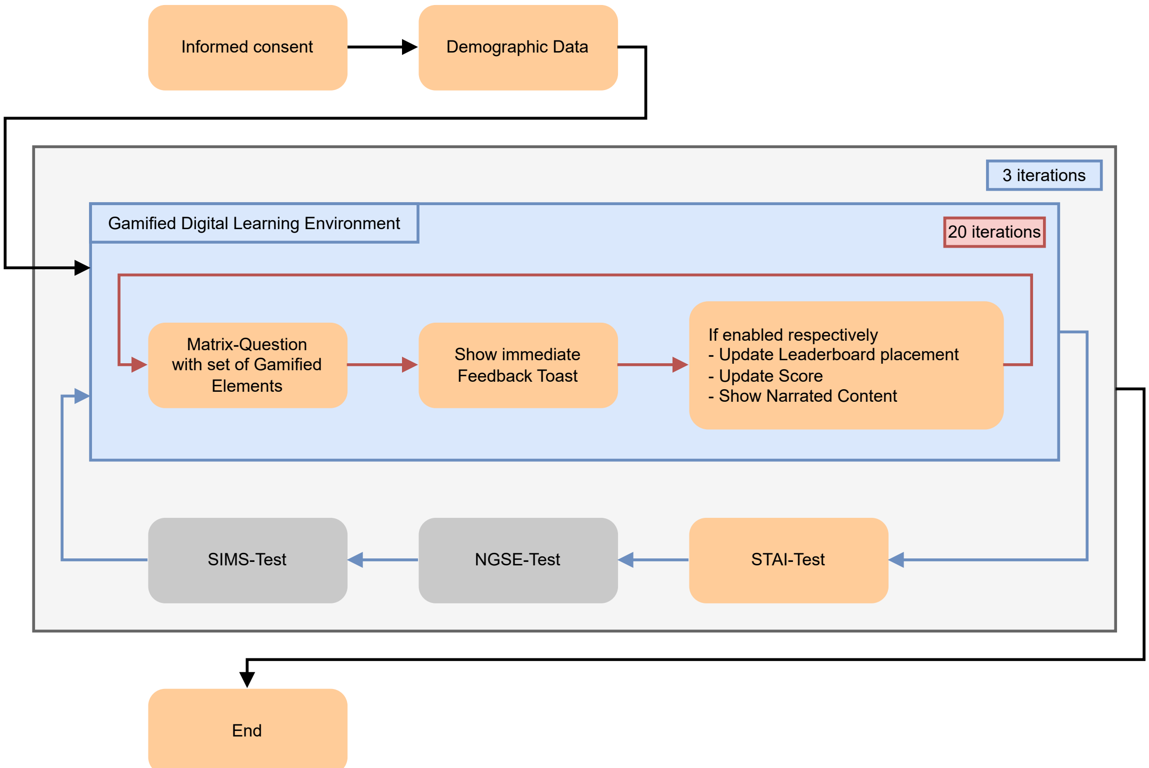


Figure 5

Flowchart depicting the study procedure. The grey area is the study without utility pages like demographic data inputs, the blue area is the gamified learning environment. The SIMS- and NGSE-tests were not part of the procedure required for this thesis.

After a brief introduction to the study's framework, participants were instructed to to proceed at their own pace and were asked to avoid any communication with other participants throughout the study to prevent data contamination. The initial interface they encountered obtained informed consent, followed by a screen to collect demographic information. After entering their demographic data, participants progressed to the core of the study which consisted of three iterations of the gamified learning environment and the questionnaires afterward. Each block began with a session in the gamified learning environment. This was followed by the completion of the three questionnaires and each concluded with the data submission screen. The transition between consecutive questions in the learning environment included a one-second delay, which extended to four seconds when narrated content was presented. This standardized sequence ensured the reliability of the measurement process and comparability across different stages of the study. After completing all three blocks, participants were thanked for their participation and compensated for their time.

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:** The performance was calculated as the ratio of correctly answered questions to the total number of questions.
- **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 one to four. The original test contains both anxiety-present and anxiety-absent items, with appropriate weights applied as per the guidelines by Court et al. (2010). Negative weights were used for items indicating anxiety (e.g., "Right now I am worried"). The initial studies by Marteau and Bekker (1992) reported a Cronbach's alpha of 0.82 for the 6-item STAI, indicating good internal consistency. Subsequent Rasch analysis revealed an item separation reliability coefficient of 0.99 and a person separation reliability coefficient of 0.78, demonstrating the scale's refined measurement precision and ability to distinguish between varying anxiety levels Court et al., 2010.

Results

Data Exclusion

Initially 120 students participated in the study. 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. Of the 120 data sets, 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 Dean and Nielsen (2007)). 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 (Nelder & Mead, 1965) 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 Corbeil and Searle (1976)). 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 (Satterthwaite, 1946) 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.

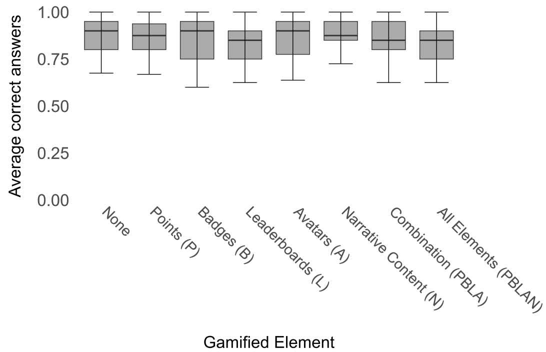
Report of findings

Performance

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. Despite some descriptive effects, our analysis revealed no significant main effects or interactions for all hypotheses regarding performance, as documented in Table 1.

Anxiety

Some descriptive trends emerged from our analysis, as illustrated in Figure 7. Women experienced higher anxiety levels than men in the non-gamified environment (M = -1.3, SE = .188 for women compared to M = -1.62, SE = .212 for men) and when



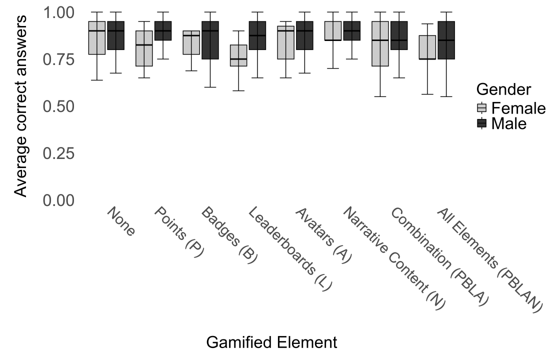


Figure 6

On the left: Overall performance across different gamification elements as percentage. On the right: Performance by gender grouped by gamification element as percentage.

leaderboards were used (*M* = -.79, *SE* = .46 for women compared to *M* = -.1.19, *SE* = .217 for men). In contrast, the use of avatars was associated with lower anxiety levels for women (*M* = -1.91, *SE* = .20) than for men (*M* = -1.4, *SE* = .177).

Despite these trends, our analysis revealed no significant main effects or interactions for any of the hypotheses regarding anxiety levels, as detailed in Table 2. This leads us to conclude that the hypotheses regarding anxiety must be rejected.

Discussion

In recent years gamification has gained popularity as a tool to enhance learner performance in various domains, including education. However, the effects of gamification elements on different groups, particularly in digital learning environments, remain underexplored. 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

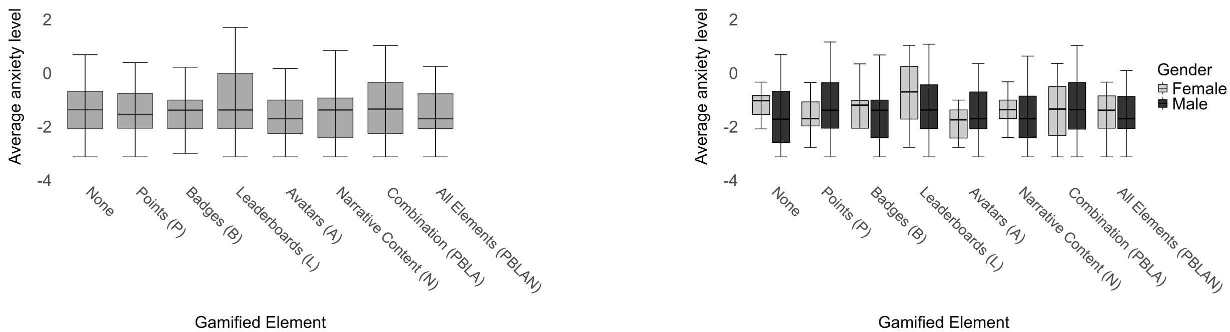


Figure 7

On the left: Anxiety levels across different gamification elements. On the right: Differences in anxiety levels by gender grouped by gamified element.

different genders react to these gamified elements regarding their cognitive and affective states.

Although the study did not find statistically significant effects, the results obtained were in line with existing research previously mentioned, indicating consistent patterns in how participants especially of different gender reacted to the gamification elements. This suggests that while the group differences were not strong enough to be statistically significant, the observed behaviours reflected known theories and previous studies.

Interpretation

This study's findings did not reveal statistically significant effects but group differences, suggesting that gamification impacts might be subtle and context-specific. These results align with prior research (Hamari et al., 2014), which also reported variable impacts of gamification on learning. The study underscores the need for cautious interpretation given the complexity of gamification's effects on different learners.

Implications

The absence of statistically significant effects in this study does not allow for definitive conclusions regarding the impact of gamification on learning outcomes. This result challenges the theoretical assumptions laid out early in this thesis, suggesting a more complex interaction between gamification and learner engagement that is likely dependent on contextual and individual differences. The observed group differences, although not significant, hint at the variability in how learners respond to gamified elements.

From a practical perspective, the non-significant findings although not suggesting specific solutions imply that the one-size-fits-all approach to gamification may not be effective for all learners. Educational software designers might consider developing more adaptive gamification strategies that are tailored to the educational and personal context of their learners. Such customization could enhance the effectiveness of educational interventions by aligning more closely with the diverse needs of students. As mentioned, without significant effects this study suggests the potential benefits of continuous adaptation and testing of gamification strategies in educational settings. Future research should aim to explore and validate these strategies as also suggested by Dehghanzadeh et al. (2024), Koivisto and Hamari (2019), and Oliveira et al. (2023), but also consider the limitations this study showed.

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 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 learning room, where officially only computer science students have access. This building is on the 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. It would have also been beneficial to allow users to personalize their avatars by entering their names, as this could further enhance user engagement and identification with the avatar.

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 additional dialogues that could enhance user interaction, while consuming additional time until the user could proceed. This might have contributed to some participants recalling to ignore the gamified elements altogether, 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 by the participants post-study was highlighted as a concern. Future studies should consider incorporating mechanisms for feedback. This could provide valuable insights into the participants' experiences and perceptions of the gamified elements, potentially revealing more nuanced effects that were not captured in this study. 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 based on our results, the gamified elements showing descriptive trends, especially leaderboards and avatars in this study should be further explored. For example, different leaderboard styles with varying avatar integrations could be tested in order to mitigate gender differences. As badges seem to have a negative outcome on anxiety, further research could explore the use of avatars in various implementations to counteract this effect for female students. 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 ITS focus on supporting the learning process through building a model for teaching in the style best suited for the individual, future studies should implement some sort of intelligence, not only in gamified elements but also the learning content shown to the student.

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 German Research Foundation (DFG; DFG reference number UP 31/1) for funding the Stuttgart Research Focus Interchange Forum for Reflecting on Intelligent Systems (IRIS). This research was further supported by the Ministry of Science, Research, and the Arts (MWK) Baden-Württemberg, Germany, within the Artificial Intelligence Software Academy (AISA; reference number Az. 7542.2-9-47.10/36/2).

Table 1

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

Variable	*beta*	*p*	*t*	*df*
m	0.07	.994	0.27	277.62
P	-0.33	.493	-1.35	211.82
B	0.06	.994	0.21	214.27
L	-0.48	.384	-1.67	213.69
A	0.13	.994	0.49	208.19
N	0.44	.384	1.70	207.03
PBLA	0.01	.994	0.06	208.37
PBLAN	0.06	.994	0.22	210.22
m × P	0.42	.493	1.33	214.33
m × B	-0.22	.994	-0.64	213.78
m × L	0.61	.384	1.79	214.82
m × A	0.01	.994	0.03	209.52
m × N	-0.00	.994	-0.01	210.19
m × PBLA	0.37	.548	1.18	211.59
m × PBLAN	0.01	.994	0.02	210.43

Note. Abbreviations: m = Male, P = Points, B = Badges, L = Level, A = Avatars, N = Narrative Content, PBLA = Combination of Points, Badges, Level, and Avatars, PBLAN = Combination of all elements. Male is compared to female performance, gamified elements are compared to the non-gamified environment. The interactions are compared to female in the non-gamified environment.

Table 2

Results of the linear mixed model analysis for STAI effects.

Variable	*beta*	*p*	*t*	*df*
m	-0.25	.623	-0.83	288.25
P	-0.28	.564	-1.00	216.76
B	0.23	.623	0.73	221.07
L	-0.04	.913	-0.11	220.38
A	-0.62	.228	-2.11	213.37
N	-0.53	.241	-1.81	211.70
PBLA	-0.03	.913	-0.11	213.53
PBLAN	-0.60	.228	-2.04	215.77
m × P	0.42	.481	1.18	220.20
m × B	-0.18	.787	-0.47	220.26
m × L	0.28	.623	0.74	221.54
m × A	0.64	.241	1.79	214.84
m × N	0.50	.437	1.40	215.47
m × PBLA	0.10	.882	0.29	217.43
m × PBLAN	0.45	.481	1.25	215.91

Note. Abbreviations: m = Male, P = Points, B = Badges, L = Leaderboards, A = Avatars, N = Narrative Content, PBLA = Combination of Points, Badges, Leaderboards, and Avatars, PBLAN = Combination of all elements. Male is compared to female anxiety, gamified elements are compared to the non-gamified environment. The interactions are compared to female in the non-gamified environment.

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

Corbeil, R. R., & Searle, S. R. (1976). Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model. *Technometrics*, *18*(1), 31–38. https://doi.org/10.1080/00401706.1976.10489397

Court, H., Greenland, K., & Margrain, T. H. (2010). Measuring Patient Anxiety in Primary Care: Rasch Analysis of the 6-item Spielberger State Anxiety Scale. *Value in Health*, *13*(6), 813–819. https://doi.org/10.1111/j.1524-4733.2010.00758.x

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

Dean, C. B., & Nielsen, J. D. (2007). Generalized linear mixed models: A review and some extensions. *Lifetime Data Analysis*, *13*(4), 497–512. https://doi.org/10.1007/s10985-007-9065-x

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

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

Lais Z Pedro, Aparecida M Z Lopes, Prates, B. G., Vassileva, J., & Isotani, S. (2015). Does Gamification Work for Boys and Girls? An Exploratory Study with a Virtual Learning Environment. https://doi.org/10.13140/RG.2.1.4783.5686

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

Lucas, K., & Sherry, J. L. (2004). Sex Differences in Video Game Play: A Communication-Based Explanation. *Communication Research*, *31*(5), 499–523. https://doi.org/10.1177/0093650204267930

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, *32*(5), 481–498. https://doi.org/10.1016/j.intell.2004.06.008

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

Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, *7*(4), 308–313. https://doi.org/10.1093/comjnl/7.4.308

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

Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin*, *2*(6), 110. https://doi.org/10.2307/3002019

Statistisches Bundesamt. (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

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

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265. https://doi.org/10.1037/a0031311

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