CBIE 2023
CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO
Uma escola para o futuro: Tecnologia e conectividade a serviço da educação

Disengaging the engaged ones: gender-stereotyped tutoring systems are undermining self-efficacy and flow state of girls

Francys Rafael do Nascimento Martins[1], Ig Ibert Bittencourt Santana Pinto[1], Geiser Chalco Challco[2], Seiji Isotani[2], Breno Felix de Sousa[1], Jessica Fernanda Silva Barbosa[1], Marcelo Reis[1]

[1]Instituto de Informática – Universidade Federal de Alagoas (UFAL)
57.072-970 – Maceió – AL – Brazil

[2]Instituto de Ciências Matemáticas e de Computação - Universidade de São Paulo (USP) - São Carlos - SP - Brazil

{francysmartins.r, ig.ibert, marceloreis.bio}@gmail.com, geiseres@ufersa.edu.br, {bfs, jfsb}@ic.ufal.br, sisotani@icme.usp.br

Abstract. *Gender Stereotyped design elements in gamification, such as using avatars aligned exclusively with the preferences of the male gender, may lead women not to develop their aptitudes or not perform adequately in male-prevalent groups, such as STEM fields (Science, Technology, Engineering, and Mathematics). This study aims to analyze the effects of stereotyped game elements in gamified tutoring systems on flow experience, self-efficacy, and learning performance. With this intent, we conducted an experimental study with Brazilian high school students. Findings indicated that stereotypes negatively affect the self-efficacy levels of female participants. Performance and flow experience levels were not affected by stereotypes. However, male participants were found to be more engaged (flow state) than women in the gamified tutoring system. These results suggest that educational technologies, such as tutoring systems and gamification, must have adaptable interfaces to avoid stereotyping and promote gender equity.*

1. Introduction

Among the various uses of technology in the classroom, one that is in development and is being widely studied is gamification. The main goals of gamification are to decrease frustration and demotivation among students, as well as improve concentration, engagement, and learning (Cózar-Gutiérrez; Sáez-Lópes, 2016; Paiva et al., 2016). Gamified educational systems comprise game elements such as trophies, levels, challenges, and leaderboards. These elements act as extrinsic motivators directly related to intrinsic motivations, such as competition, self-expression, accomplishment, status, reward, and selflessness. This leads to an increase in the motivation and engagement of students (Liu, 2020).

Despite the various benefits of gamification, on some occasions, this approach may not provide the necessary support to the student and lead to adverse effects (Almeida; Kalinowski e Feijó, 2021). This happens when the design is inadequate. This study pertains to the use of the inadequate sexist design of visual elements and game mechanics. When the design is not aligned with the participant's gender, this constitutes a stereotype threat, which may harm psychological indicators, leading to discomfort, lack of motivation, low self-efficacy, and performance decrease (Pennington et al., 2016).

Gamified design that disagrees with a participant's gender is a phenomenon that has been previously studied by Albuquerque et al. (2017), who found that male stereotypes increase women's anxiety levels. Likewise, we hypothesize that individual factors, such as self-efficacy and flow state, may also be affected by gender stereotypes. Self-efficacy may be harmed by gender stereotypes because game elements, such as a male-dominated ranking, may decrease one's belief in their own abilities to execute a particular task. Flow state concerns the mental state of complete immersion in an activity (Csikszentmihalyi, 1990). This is a desired mental state within gamified educational settings that may be hindered by gender stereotypes, leading to lower learning performance. For example, a color design in disagreement with the participant's gender may cause discomfort and prevent the student from reaching a flow state.

Based on the premises above, we conducted a quasi-experiment to investigate the influence of gender stereotypes on students' self-efficacy, flow state, and performance levels. We implemented a gamified tutoring system in three versions: male-stereotyped (StMale), female-stereotyped (stFemale), and non-stereotyped (Default).

2. Related Studies

Until now, only two studies in the scientific literature have approached the effects of stereotype threat in gamified learning environments.

Christy and Fox (2014) conducted a study in which female undergraduate students interacted with three versions of a virtual representation of a classroom: in the first, the top-ranking participants were only men; the second version had women as the top-ranking participants; the third had no ranking at all. The results of the study showed that when women were assigned to a female-dominated environment (stereotypes aligned with their gender), they presented lower performances compared to participants who were assigned to a male-dominated setting (stereotype threat). Therefore, the study presented evidence that stereotype threat may lead women to reject the stereotypes. This phenomenon is known as stereotype boost.

Albuquerque et al. (2017) developed an experiment to investigate whether gender stereotype threat in gamified educational scenarios influences anxiety and performance. The authors designed a study with two anxiety tests (a pre-test and a post-test) and prototypes of a male-stereotyped gamified tutoring system, a female-stereotyped one, and a non-stereotyped system. The participants were Brazilians between 18 and 54 (82 men and 45 women). The findings indicated that men and women had different anxiety levels when faced with stereotype threats. Women in the male environment had significant increases in anxiety levels, which could harm learning performance.

The studies of Christy and Fox (2014) and Albuquerque et al. (2017) showcase distinct effects that stereotype threat may produce. A plausible explanation for these differences may be that the effects of stereotypes vary according to age (Miller et al., 2018). The first study comprised students above 18, while the second had participants of any age. In this context, our study aims to provide insight into another target audience:

adolescents. Also, the two studies mentioned above did not assess the effects of stereotypes on self-efficacy, which is the main objective of this study.

3. Methodology

The study was conducted as a pre-test/post-test quasi-experiment in a 2x3 factorial design, as illustrated in Figure 1. The first factor was composed of two conditions: female participants and male participants. The participants were distributed among three groups (Group I, Group II, and Control Group). The second factor comprised three conditions, which corresponded to the three versions of the gamified platform: male-stereotyped (stMale), female-stereotyped (stFemale), and non-stereotyped (Default).

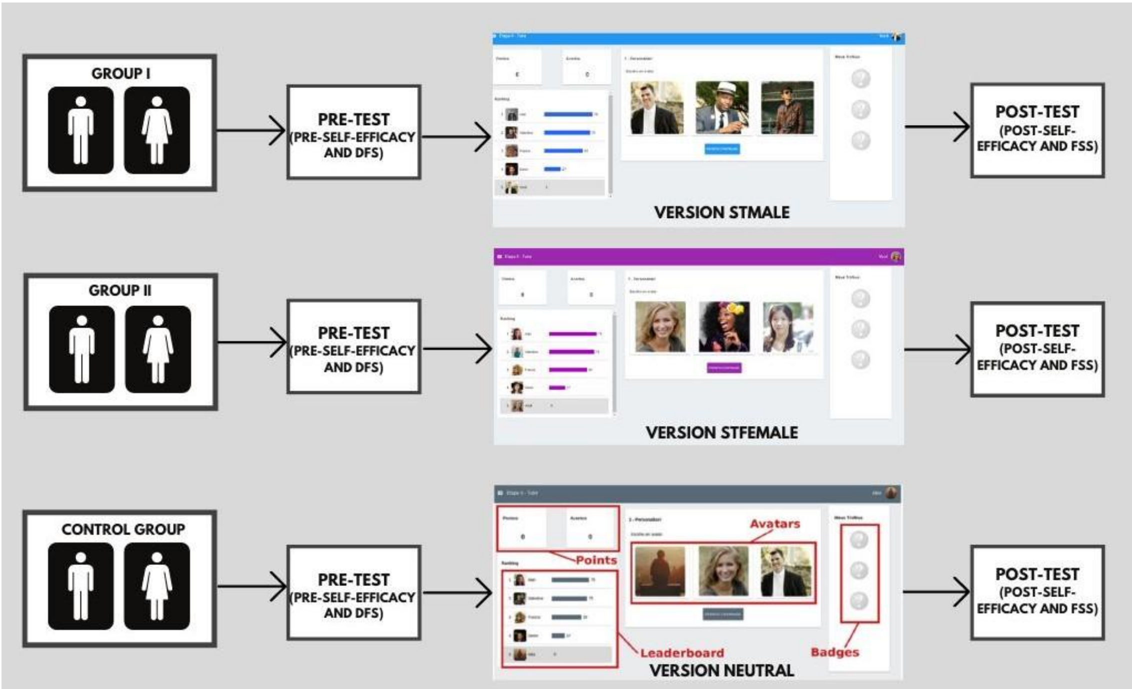

Figure 1. Study overview.

3.1. Formulation of hypotheses

The experimental design was implemented to assess the following null hypotheses.

H1: There is no significant difference in the self-efficacy levels of participants regardless of the condition they were assigned (stThreat - when the stereotype disagrees with the participant's gender; stBoost - when the stereotype agrees with the participant's gender; and neutral - when there are no stereotypes).

H2: There is no significant difference in the self-efficacy levels of participants regardless of their gender (male and female) and the condition they were assigned (stMale, stFemale, and Default).

H3: There is no significant difference in the flow state of participants regardless of the condition they were assigned (stMale, stFemale, and Default).

H4: There is no significant difference in the flow state of participants regardless of their gender (male and female) and the condition they were assigned (stMale, stFemale, and Default).

H5: There is no significant difference in participants' performances regardless of the condition they were assigned (stMale, stFemale, and Default).

H6: There is no significant difference in participants' performances regardless of their gender (male and female) and the condition they were assigned (stMale, stFemale, and Default).

3.2. Measures and covariates

As seen in Figure 1, the study was conducted in three steps: pre-test, execution, and post-test. This allowed us to compare self-efficacy levels before (pre-test) and after (post-test) using the three versions of the gamified platform (stMale, stFemale, and Default). As a measure of self-efficacy, a six-item questionnaire was applied for both the pre-test and post-test phases. Each item consisted of a mathematics question (similar to the exercises in the execution phase).

After solving each question, the participants were asked to answer, on an eight-point Likert scale, how effective they felt in the task.

The questionnaire of DFS-2 was applied to measure the dispositional flow state, and the FSS-2 was used to measure the flow state. The scales were used in both, the pre-test and post-test phases, and in their short and Brazilian Portuguese-adapted versions (Bittencourt et al., 2021).

Participants' learning performances were measured with the scores obtained in the gamified platforms called ActivityPoints. The points were added to the user's score for each correct answer.

3.3. Manipulations or experimental interventions

As demonstrated in Figure 1, the three versions of the system had the same interface and gamified elements, such as PBL (Points - Badges - Leader boards) and avatars. In every version, the participant began with zero points and obtained five by choosing an avatar. By answering correctly, the participant won five points, while no penalty was given for an incorrect answer. When achieving 25 and 50 points, the student won a badge that informed him of the respective achievement. The activity ended when the participant answered 20 questions.

In the non-stereotyped setting, the interface and game elements were gray, the user could choose from a male, female or androgynous avatar, and the ranking presented users of both genders (female and male). In the male-stereotyped setting, the predominant color was blue, the avatars were male, and the leaderboard was composed exclusively of men. As for the female-stereotyped versions, the color gradient was mainly lilac, the avatars were female, and the ranking was made up of females.

3.4. Sampling procedure and characteristics of participants

In this study, participant sampling was random. The students participated voluntarily in the experiment and were invited through Whatsapp groups of the state schools Comendador José da Silva Peixoto, Ernani Méro and Alcides Andrade, and the Federal Institute of Alagoas in the city of Penedo. The teachers helped in the experiment's application, as it occurred during class and was assigned as complementary coursework.

The participants were (n=82) high school students: 42.7% (n=35) men and 57.3% (n=47) women. As for race, 65.9% (n=54) were Brown, 17.1% (n=14) were White, 9.8% (n=8) were Black, 3.7% (n=3) were Asian, and 3.7% (n=3) chose not to declare their race. Concerning economic status, 52.4% (n=43) were middle-class students, 43.9% (n=36)) were lower-class, and 3.7% (n=3) were upper-class.

3.5. Data collection process

In conformity with the design specified in Figure 1, each participant was randomly assigned to one of the three groups, out of which two corresponded to the stereotyped conditions (stMale and stFemale) and one was the control group (non-stereotyped). The data collection was performed from May 2021 to August 2021. A superficial statement was elaborated with the basic instructions to access and use the platform. Details of the study were not provided.

During the data collection process, the researchers had no contact with the participants, seeing that the teachers demanded that they apply the experiment by sending the students the link to the platform. In the pre-test, each participant answered the self-efficacy and the DFS-2 questionnaires. During the execution phase, participants tried to solve the 20 mathematics questions. Lastly, the participants answered the self-efficacy and FSS-2 questionnaires in the post-test. In this last phase, the participants also filled out a socioeconomic questionnaire on race, gender, and socioeconomic status.

3.6. Statistics and data analysis

The statistical parametric ANCOVA and ANOVA tests were applied to assess significant differences in flow state, self-efficacy, and learning performance levels. Before running these tests, to reduce the effect of outliers with extreme values, we applied the winsorized method with a 5% to 95% probability within the collected data. Furthermore, we used the Shapiro-Wilk test to assess whether the symmetry and normality conjectures were satisfactory. Every test was performed with the R studio version 4.1.0 (R Core Team, 2021) and the R package “rshinystatistics” version 0.0.1.

4. Results and Discussion

Table 1 presents the descriptive statistics of the collected data, which are organized in the following manner: the main columns contain the data according to condition (stThreat, stBoost, and neutral) and environment (stMale, stFemale, and default); the rows present participants' data for pre-test (pre.self) and post-test (pos.self) self-efficacy, dispositional flow state (dfs), flow state (fss), and scores (points). The table also presents the adjusted self-efficacy (adj.self) and flow state scale (adj.fss) values through Estimated Marginal Means (EMMs).

Table 1. Descriptive data of self-efficacy, disposition to flow state (dfs), flow state scale (fss), points, and their adjusted values (adj.self, adj.fss) trhough EMMs.

		Condition			stMale		stFemale		default	
		stThreat	**stBoost**	**neutral**	**male**	**female**	**male**	**female**	**male**	**female**
pre.self	**N**	28	23	27	11	17	11	14	13	16
	M	5.838	5.854	6.067	5.902	5.129	6.848	6.122	6.572	5.841
	SE	259	376	289	632	299	262	437	397	386
pos.self	**N**	28	23	27	11	17	11	14	13	16
	M	5.643	5.952	5.919	6.539	5.106	6.386	5.845	6.629	5.617
	SE	223	344	258	468	0.28	286	443	299	338
adj.self	**M**	5.687	5.988	5.842	6.589	5.549	5.955	5.783	6.338	5.698
	SE	213	235	217	325	271	333	288	303	0.27
dfs	**N**	27	25	27	11	14	11	14	13	16
	M	3.495	3.803	3.659	3.869	3.556	3.449	3.744	3.701	3.701
	SE	88	88	104	131	154	86	113	193	0.1
fss	**N**	27	25	27	11	14	11	14	13	16
	M	3.396	3.675	3.733	4.015	3.333	3.697	3.419	3.86	3.616
	SE	139	163	148	187	181	181	222	234	187
adj.fss	**M**	3.487	3.583	3.727	3.891	3.406	3.837	3.373	3.841	3.597
	SE	140	146	138	205	180	205	180	186	168
points	**N**	23	21	29	11	17	11	14	13	16
	M	50.217	51.667	52.931	55.000	44.118	41.818	38.571	54.615	51.562
	SE	23.085	25.658	20.068	7.362	5.922	7.362	6.525	6.772	6.104

In Table 2, we describe the paired comparisons of the distinct conditions, genders, and environments, performed after the ANCOVA and ANOVA tests, as well as the results for the hypotheses regarding self-efficacy, flow state, and learning performance.

Table 2. Pairwise comparisons and hypothesis tests for self-efficacy (H1, H2), flow state (H3, H4), and learning performance (H5, H6).

	var	**(I)**	**(J)**	**(I)-(J)**	**SE**	**Stats**	**p**	**p.adj**	**Decision**
H1	Condition	stBoost	stThreat	0.301	0.317	0.950	0.345	1	Fail to reject
		stBoost	neutral	0.146	0.320	0.457	0.649	1	
		stThreat	neutral	-0.155	0.304	-0.509	0.613	1	
H2	Gndr:female	default	stFemale	-0.085	0.395	-0.215	0.831	1	Reject
	Gndr:female	default	stMale	0.149	0.380	0.392	0.696	1	

	Gndr:female	stFemale	stMale	0.234	0.398	0.588	0.558	1	
	Gndr:male	default	stFemale	0.383	0.442	0.866	0.389	1	
	Gndr:male	default	stMale	-0.251	0.445	-0.564	0.574	1	
	Gndr:male	stFemale	stMale	-0.634	0.467	-1.360	0.178	0.534	
	env:default	female	male	-0.640	0.407	-1.570	0.121	0.121	
	env:stFemale	female	male	-0.171	0.439	-0.391	0.697	0.697	
	env:stMale	female	male	-1.040	0.422	-2.462	0.016	0.016	
H3	Condition	stBoost	stThreat	0.096	0.206	0.467	0.642	1	Fail to reject
		stBoost	neutral	-0.144	0.201	-0.716	0.476	1	
		stThreat	neutral	-0.240	0.197	-1.216	0.228	0.684	
H4	Gndr:female	default	stFemale	0.224	0.246	0.912	0.365	1	Reject
	Gndr:female	default	stMale	0.191	0.247	0.775	0.441	1	
	Gndr:female	stFemale	stMale	-0.033	0.255	-0.129	0.898	1	
	Gndr:male	default	stFemale	0.005	0.278	0.016	0.987	1	
	Gndr:male	default	stMale	-0.049	0.276	-0.179	0.858	1	
	Gndr:male	stFemale	stMale	-0.054	0.294	-0.184	0.855	1	
	Env: default	female	male	-0.244	0.251	-0.974	0.333	0.333	
	Env: stFemale	female	male	-0.464	0.274	-1.689	0.096	0.096	
	Env: stMale	female	male	-0.485	0.275	-1.762	0.082	0.082	
H5	Condition	stBoost	stThreat	1.449	6.861	0.211	0.833	1	Fail to reject
		stBoost	neutro	-1.264	6.513	-0.194	0.847	1	
		stThreat	neutro	-2.714	6.347	-0.428	0.670	1	
H6	Gndr:female	default	stFemale	12.991	8.935	1.454	0.150	0.45	Fail to reject
	Gndr:female	default	stMale	7.445	8.504	0.875	0.384	1	
	Gndr:female	stFemale	stMale	-5.546	8.812	-0.629	0.531	1	
	Gndr:male	default	stFemale	12.797	10.002	1.279	0.205	0.614	

	Gndr:male	default	stMale	-0.385	10.002	-0.038	0.969	1	
	Gndr:male	stFemale	stMale	-13.182	10.411	-1.266	0.209	0.628	
	Env: default	female	male	-3.053	9.117	-0.335	0.739	0.739	
	Env: stFemale	female	male	-3.247	9.837	-0.330	0.742	0.742	
	Env: stMale	female	male	-10.882	9.448	-1.152	0.253	0.253	

4.1. Self-efficacy per stereotype condition (H1)

Even though students' self-efficacy levels increased when in the stBoost condition and decreased in the other two (stThreat and neutral), we did not reject the null hypothesis H1. According to the results indicated in Table 2, no significant differences in self-efficacy according to stereotype condition were seen.

The self-efficacy indicators found in this study are aligned with the results of the study of Spencer, Steele and Quinn (1999), who found that self-efficacy was not affected by the reducing stereotype threat. This reduction is comparable to the stBoost and neutral conditions in our study, considering that the participants were not exposed to disagreeing stereotypes in both conditions.

Our findings diverge from those of Chung et al. (2010), who found that the perception of threat results in lower self-efficacy levels. However, it is worth considering that our study was conducted with high school students, while that of Chung et al. (2010) comprised college students. Also, this work implemented a different assessment method, which consisted of measuring pre-test and post-test results with quantitative psychometric instruments, unlike Chung et al. (2010), who used a written test to assess stereotype threat These elements may have led to the disagreeing results.

4.2. Self-efficacy per gender and environment (H2)

Based on the statistical results presented in Table 2, we reject null hypothesis H2. When women participated in a male-stereotype environment, their self-efficacy levels were statistically significantly lower than men's. These findings indicated that male-stereotyped gamified platforms might have a negative impact on women's self-confidence. Therefore, it suggests that the male-stereotyped settings may contribute to gender inequity and make women feel they do not belong in these male-dominated environments, such as the STEM fields. This may occur because male stereotypes can hinder women's belief in their ability to perform tasks.

Regarding null hypothesis H2, our results are similar to those of Mayer and Hangs (2003), whose study showed that the group under specific stereotype threat had their self-efficacy levels negatively impacted. In the study, Black individuals experienced stereotype threat and, as a result, presented lower self-efficacy levels compared to White individuals. Our results also agree with those of Chung et al. (2010), who saw a relationship between self-efficacy and anxiety caused by increased threat perception. Thus, it is reasonable to suggest that the same may have occurred with the women in our study; that is, they felt more anxious due to the male stereotype threat, leading to a decrease in their self-efficacy.

4.3. Flow experience per condition (H3)

Table 2 shows no statistically significant differences in flow state per stereotype condition were verified. Thus, participants' flow states were not altered by the stBoost, stThreat, and neutral conditions. Hence, null hypothesis H3 was accepted. Nevertheless, it is essential to highlight that the dispositional flow (dfs) of participants in the pre-test was higher in the neutral condition compared to the other two. This suggests that the threat condition did not induce significant adverse effects on flow state (fss), as expected initially.

We consider that our findings may complement those of Simões et al. (2015), Li (2019), and Cheng, Su and Kunshuk (2015), who found that gamification positively affects flow state. In this sense, our results suggest that the benefits of gamification for flow state could possibly not be affected by stereotype threat.

4.4. Flow experience per gender and environment (H4)

We reject null hypothesis H4 because a statistically significant difference in the flow state scale (fss) was seen among participants of different genders. This means that despite the type of gender stereotype, male students had higher flow state levels. Thus, engagement promotion through gamification was more effective for men.

The evidence in our study, which indicates a negative relation between flow state and anxiety, disagrees with Albuquerque et al. (2017), who found that women significantly increased performance due to gender stereotypes.

4.5. Performance and learning per condition (H5)

The statistics show a failure to reject the null hypothesis H5, given that there were no statistically significant differences in learning performance measured with ActivityPoints. Regardless of the stereotype condition, no learning performance differences were seen.

4.6. Performance per gender and environment (H6)

The results indicate no statistically significant differences in participants' performances per gender and type of stereotyped environment. Thus, null hypothesis H6 is accepted. Men and women had similar scores in any of the three environments regardless of gender.

The gender stereotypes in the three gamified versions did not affect academic performance. These findings are opposed to those of Spencer, Steele and Quinn (1999), and Mayer and Hangs (2003), whose results suggested that stereotypes harmed performance. Nevertheless, we stress that these studies used explicit gender stereotypes, such as direct messages (e.g. "boys are better than girls in mathematics"). As for our study, we applied implicit gender stereotypes related to the assigned tasks and the interfaces and game elements, which may affect motivation, engagement, and performance.

5. Conclusion and Future Studies

From the data obtained in this study, it is possible to conclude that in gamified tutoring systems, male-stereotyped elements can harm women's self-efficacy levels. Moreover, these stereotypes do not affect flow experience or academic performance. Concerning flow experience, we found statistically significant effects conditioned by gender. Women were less immersed than men. However, this was not attributed to the gender stereotypes in the setting. Future studies should be performed to better explain these findings. For example, it may be relevant to investigate whether the proposed tasks or another factor may lead to this significant difference.

The results point to the need for better elaborating and improving gamified digital technology to avoid gender stereotyping and promote gender equity to ensure satisfactory learning performances and avoid disengagement, as seen in this study. In this sense, our study provides insight into various Artificial Intelligence (AI) techniques, such as Computational Vision, Pattern Recognition, and Natural Language Processing, to identify gender stereotypes. An example is the study of Santos et al. (2022), in which female and male color biases within educational web technologies were calculated. Furthermore, our study reinforces the need for adaptive interface mechanisms to better align with participants and avoid stereotype threats. Such mechanisms may be developed with the use of AI tools.

Also, our findings hint at the need for future studies. Particularly to understand why women's self-efficacy levels were affected when male-stereotyped game elements were included in the system. We also point to the need to better comprehend the effects that this self-efficacy decrease may induce. However, it is worth noting that the self-efficacy results did not affect performance, as seen in the study of Rafiola et al. (2020). Or even if self-efficacy did affect performance to some degree, it was not enough to produce significant results, such as in the studies of Shin (2018) and Alghamdi et al. (2020).

We perceive that the rate of students of completed the experiment is a limitation of this work. This number was considerably inferior to the number of students who received the invitation. Because this experiment was conducted during the Covid-19 pandemic, some students reported being tired of online learning. Therefore, the results could have been different if the experiment had been conducted in person.

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. In *Computers & Education*, pages 161–170. Publishing Press.

Alghamdi, A., Karpinski, A.C., Lepp, A., Barkley, J., (2020). Online and face-to-face classroom multitasking and academic performance: Moderated mediation with self-efficacy for self-regulated learning and gender. In *Computers in Human Behavior*, pages 214–222. Publishing Press.

Almeida, C., Kalinowski, M., Feijó, B., (2021). A systematic mapping of negative effects of gamification in education/learning systems. In *2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pages 17–24.

Bittencourt, I.I., Freires, L., Lu, Y., Challco, G.C., Fernandes, S., Coelho, J., Costa, J., Pian, Y., Marinho, A., Isotani, S., (2021). “Validation and psychometric properties of the brazilian-portuguese dispositional flow scale 2 (dfs-br)”. In *PloS one*, https://doi.org/10.1371/journal.pone.0253044

Cheng, M.T., Lin, Y.W., She, H.C., (2015). Learning through playing virtual age: Exploring the interactions among student concept learning, gaming performance, in-game behaviors, and the use of in-game characters. In *Computers & Education*, pages 18–29. Publishing Paper.

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. In *Computers & Education*, pages 66–77. Publishing Paper.

Chung, B.G., Ehrhart, M.G., Holcombe Ehrhart, K., Hattrup, K., Solamon, J., (2010). Stereotype threat, state anxiety, and specific self-efficacy as predictors of promotion exam performance. In *Group & organization management*, pages 77–107. Publishing Paper.

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

Li, C.H., (2019). Gamification of an asynchronous html5-related competency-based guided learning system. In IOP Conference Series: Materials Science and Engineering, page 12004. Publishing Paper.

Liu, S., (2020). Developing a gamified online learning management system for business students.

Mayer, D.M., Hanges, P.J., (2003). Understanding the stereotype threat effect with" culture-free" tests: An examination of its mediators and measurement. In *Human Performance*, pages 207–230. Publishing Paper.

Miller, D.I., Nolla, K.M., Eagly, A.H., Uttal, D.H., (2018). The development of children’s gender-science stereotypes: A meta-analysis of 5 decades of us draw-a-scientist studies. In *Child development*, pages 1943–1955. Publishing Paper.

Paiva, R., Bittencourt, I.I., Tenório, T., Jaques, P., Isotani, S., (2016). What do students do on-line? modeling students’ interactions to improve their learning experience. In *Computers in Human Behavior*, pages 769–781. Publishing Paper.

Pennington, C.R., Heim, D., Levy, A.R., Larkin, D.T., (2016). Twenty years of stereotype threat research: A review of psychological mediators. In *PloS one*. Publishing Paper.

Rafiola, R., Setyosari, P., Radjah, C., Ramli, M., (2020). The effect of learning motivation, self-efficacy, and blended learning on students’ achievement in the industrial revolution 4.0. In *International Journal of Emerging Technologies in Learning*, pages 71–82, Publishing Paper,

Santos, J., Bittencourt, I., Reis, M., Chalco, G., Isotani, S., (2022) Two billion registered students affected by stereotyped educational environments: an analysis of gender-based color bias. In *Humanities and Social Sciences Communications*, pages 1–16. Publishing Paper.

Shin, M.H., (2018). Effects of project-based learning on students' motivation and selfefficacy. In *English Teaching*, pages 95–114. Publishing Paper.

Simões, J., Mateus, S., Redondo, R., Vilas, A.F., (2015). An experiment to assess students' engagement in a gamified social learning environment. In *eLearn*, pages 38–42.

Spencer, S.J., Steele, C.M., Quinn, D.M., (1999). "Stereotype threat and women's math performance". In *Journal of Experimental Social Psychology*, pages 4–28. https://doi.org/https://doi.org/10.1006/jesp.1998.1373

Spencer, S.J., Steele, C.M., Quinn, D.M., (1999) Stereotype threat and women's math performance. In *Journal of experimental social psychology*, pages 4–28. Publishing Paper