Journal of the Brazilian Computer Society, 2024, 30:1, doi: 10.5753/jbcs.2024.3600

When boosting gender stereotypes increases flow experience and reduces self-handicapping in gamified tutoring systems

João Vitor Lourenço Batista do Nascimento [**Federal University of Alagoas** | ***joao.nascimento@ip.ufal.br***]
Jário José dos Santos Júnior [**Federal University of Alagoas** | ***jario.junior@penedo.ufal.br***]
Geiser Chalco Challco [**University of São Paulo** | ***geiser.gcc@gmail.com***]
Ig Ibert Bittencourt [**Federal University of Alagoas** | ***ig.ibert@gmail.com***]

Institute of Computing, Federal University of Alagoas (UFAL), Campus A.C. Simões, Cidade Universitária, Maceió, AL, 57072-970, Brazil.

Received:4 August 2023 • **Accepted:**25 March 2024 • **Published:** 15 September 2024

Abstract The threat of stereotypes affects various psychological mechanisms, including affective/subjective, cognitive, and motivational ones, and can be present in gamified online educational environments in various ways. In this study, we aimed to investigate whether gender stereotypes in gamified virtual environments could affect the flow experience, self-handicapping behavior, and performance of Brazilian students. To achieve this, we experimented with 147 participants (60 males and 87 females) who were high school and higher education students from public and private institutions in the state of Alagoas, located in the northeastern region of Brazil. We randomly allocated the participants to three distinct virtual environments: a neutral environment, a male-stereotyped environment, and a female-stereotyped environment. We introduced the stereotype threat condition when the participant was in an environment that did not correspond to their gender. In contrast, the boost condition occurred when the environment corresponded to their gender. The results of this study indicate that the presence of gender stereotypes can influence both the flow experience and the self-handicapping behavior of Brazilian students in gamified virtual environments. We observed statistically significant differences that suggest that the implementation of stereotypes can influence the relationship between variables.

Keywords: Computer Science in Education, Flow, Gamified Virtual Environments, Gender Stereotypes, Self-handicapping.

1 Introduction

In recent decades, stereotype threat has garnered significant attention due to its potential to adversely affect social interactions and formal education [Steele and Aronson, 1995; Pennington *et al*., 2016]. Conversely, gamification emerged as a novel approach in the early 21st century, aiming to enhance motivation and engagement by integrating game elements into non-game contexts [Kapp, 2012].

Amidst concerns over the detrimental impact of gender stereotypes on psychological factors and learning outcomes [Pennington *et al*., 2016; Santos *et al*., 2022a,b], our inquiry delves into whether such stereotypes persist within gamified virtual environments. This investigation is crucial, particularly considering the influence of gender stereotypes on the career choices and educational pursuits of young women, notably in STEM fields where male dominance prevails [Beede *et al*., 2011; Christy and Fox, 2014; Piatek-Jimenez *et al*., 2018; Chang *et al*., 2019; Santos *et al*., 2022a]. Moreover, the underrepresentation of women in STEM is viewed as a hindrance to innovation, limiting the diversity of perspectives and exacerbating wage and employability disparities between genders [Beede *et al*., 2011; Santos *et al*., 2022b].

Albuquerque *et al*. [2017] explored the impact of gender stereotypes on anxiety levels among Brazilians within various iterations of a stereotyped gamified virtual environment, shedding light on its implications for learning. Their findings revealed that exposure to a male-stereotyped environment heightened participants' anxiety levels, consequently affecting task performance.

Drawing on the Stereotype Boost Theory [Shih *et al*., 2012] and the Multiple Threat Model of Stereotypes [Shapiro and Neuberg, 2007], our study poses the question: "How do Gender Stereotype Boost and Threat influence the flow experience, Self-Handicapping, and performance of Brazilian Students in gamified tutoring systems?" To address this, we engaged high school and undergraduate students from diverse public and private institutions in Alagoas, Northeast Brazil.

Our investigation aims to assess the potential positive effects of stereotypes on psychological factors (Flow and Self-handicapping) that may impact learning within different versions of a gamified virtual environment featuring positive gender-based feedback.

Our findings underscore the significant impact of gender stereotypes on learning mediators, manifesting in fluctuations in self-handicapping and flow levels depending on the environment. Particularly concerning is the pronounced increase in negative mediators experienced by women, reflecting the broader issue of their underrepresentation and its ramifications on psychological learning processes.

By shedding light on the incidence of gender stereotypes in learning mediators, our study seeks to contribute to the

fields of psychology and applied computer science in education. We aim to advocate for more equitable information solutions that account for the pervasive influence of gender stereotypes.

This article is structured as follows: Section 2 provides the theoretical background and related works. Section 3 presents the proposal and the tool used in this study. In Section 4, we depict the results obtained from the conducted experiment. In Section 5, we develop a discussion about the results previously reported. In Section 7, we discuss the limitations of our study. In Section 8, some future works are outlined. Finally, in Section 9, we provide our concluding remarks.

2 Background

This section presents the main concepts of the proposed work. In addition, we will depict related works to this study, more precisely related to gamified educational environments and stereotype threats in virtual environments.

2.1 Gamification and Gamified Educational Environments

According to Costa and Marchiori [2015], gamification presents from different perspectives, both as a "fad" for those in the IT (Information Technology) field and as a possible solution to different organizational and learning problems. As we know, gamification is the application of game elements in non-game contexts (non-game activities) [Kapp, 2012; Costa and Marchiori, 2015]. Its elements comprise three main categories: mechanics, dynamics, and components [Moreira and F., 2012]. It is important to note that it is essential for these elements to be effectively combined to achieve the objective of gamification. It is not enough to introduce the elements into the context [Costa and Marchiori, 2015]. Gamified online environments consist of educational tools to promote motivation and engagement in the virtual environment. We currently have several examples of platforms that use this "gamified" modality, such as Duolingo, and Wize Up, among others, that focus on promoting a more dynamic and engaged learning experience, bringing novelty and enthusiasm to users of different information systems. The work of Moreira and F. [2018] provides us with examples of elements used by Duolingo developers to implement gamification strategies to provide a better user experience. Additionally, Kim and Castelli [2021] indicates that gamification can promote positive behavior changes in favor of learning depending on how it is employed.

2.2 Stereotype Threat Theory (STT) and Stereotype Boost (SBT)

According to Myers [2014], stereotype threat is an individual's excessive apprehension that they will be evaluated based on a negative stereotype about themselves or the group to which they belong. According to Pennington *et al*. [2016], stereotype threat can shed light on quantitative intellectual differences between different genders, social classes, and ethnicities during the performance of the same test. The literature suggests that individuals who are more aware of the stereotypes surrounding themselves and their group, and those with more identification with it, tend to be more prone to suffering from this problem. It is worth noting that not all individuals in the same group experience stereotype threat in the same situations and in the same way [Shapiro and Neuberg, 2007]. As we will see below, while a stereotype may be harmful to some, it may be a factor that boosts performance for others [Shih *et al*., 2012]. According to Shih *et al*. [2012], the Stereotype Boost (SBT) theory develops in parallel with the Stereotype Threat (STT) theory. While the theory of stereotype threat concerns, at first analysis, the effects of negative stereotypes on performance, the Stereotype Boost theory analyzes how positive stereotypes can increase (or improve or implement) performance. The performance increase by stereotype occurs only in specific situations. If the conditions are not met, the increase does not occur, or it can even cause a drop in performance. The method of activating positive stereotypes is essential, as well as the characteristics of the individual reported in the stereotype, as it can or cannot have sound activation or results [Shih *et al*., 2012].

2.3 Stereotype Boost – Positive stereotypes leading to performance gains

The studies by Shih *et al*. [2012], Smith and Johnson [2006], and Swift *et al*. [2013] present different approaches to addressing positive stereotypes. In Shih *et al*. [2002], a study focused on the relevance of the individual associated with the way positive stereotypes were activated. As a result, Shih *et al*. [2002] found evidence that participants who participated in situations where positive stereotypes were activated in a non-explicit manner achieved better results, even though the stereotypes were positive. Smith and Johnson [2006] found a similar result in their study. When conducting an experiment that aimed to analyze the effect of positive stereotypes on performance and motivation in situations where individuals identified with the studied domain, the researchers found a result suggesting that positive stereotypes, when structured by comparing two groups (favoring one group over another), could lead to discomfort and even produce a state of heightened expectation, causing individuals to feel pressured to live up to the belief about their group and ultimately resulting in a decrease in performance. Swift *et al*. [2013] conducted an experiment that involved a social comparison between younger and older individuals performing tasks. In their study, the researchers concluded that different combinations of activities and comparisons could generate both stereotype threat and increased performance. The authors draw attention to two possible ways to enhance the performance of older individuals engaging in activities. The first method directly addresses any anxiety, while the other entails identifying positive expectations in positively stereotyped domains, thereby promoting more positive feedback.

2.4 Flow Experience

The state of optimal experience (flow state) has been observed across different cultures and activities throughout time. From reports of painters, writers, and musicians who

entered a deep state of concentration in their compositions to athletes and religious practitioners who, during their practices, felt disconnected from themselves and the demands of the surrounding world [Snyder *et al*., 2009]. Csikszentmihalyi [2020] characterizes Flow as a state of constant enjoyment, high concentration on the task, and promoter of intrinsic motivation. The activity must provide at least one of its eight main components to achieve the Flow state. These elements emerged in interviews conducted by him, Csikszentmihalyi [2020], during his research with diverse populations and occupations, who reported experiencing at least one, if not all, of the following elements: the first being ”balance between challenge and skill,” the second is ”merging of action and awareness,” the third and fourth consist of ”clear goals and feedback,” the fifth is ”total concentration on the task at hand,” the sixth is the ”paradox of control,” the seventh is ”loss of self-awareness.” The eighth is “transformation of time” [Csikszentmihalyi, 2020]).

2.5 Self-handicapping

Translated from Portuguese by Zanatto [2007] and also known in Brazil as “Self-Defeating Strategies,” Self-Handicapping is a set of self-protective strategies to avoid individual responsibility for failure. According to the literature review conducted by Mena [2019], some authors investigating this phenomenon have concluded that specific individuals may anticipate failure during task performance and intentionally produce situations for which they can direct blame their failure. Mena [2019] suggests two possible self-handicapping strategies: Active and Claimed. In active self-handicapping, individuals use strategies that effectively fail, whereas, in claimed self-handicapping, the strategies do not necessarily lead to failure but can be used by the individual to avoid responsibility. Moreover, active strategies can take two forms: internal, related to the individual’s harmful actions, and external, related to things “outside the realm” of the self, such as unattainable goals or external factors that interfere with concentration [Mena, 2019]. According to Török *et al*. [2018], in educational contexts, stereotypes and stigmatization may be related to changes in levels of self-handicapping. Exposure to stereotypes could lead to excessive apprehension, causing individuals to anticipate their failure and create strategies to avoid responsibility for it.

2.6 The presence of stereotypes in digital educational technologies and their effects

Digital educational technologies face challenges from conception to final use, including the presence of gender and race stereotypes. Recent studies, such as those by Christy and Fox [2014], Albuquerque *et al*. [2017], Chang *et al*. [2019], and Santos *et al*. [2022a], highlight how these stereotypes can impair performance and increase anxiety in online educational environments. For instance, Christy and Fox [2014] demonstrated that women exposed to gender stereotypes performed worse on mathematics tests in virtual reality, while Albuquerque *et al*. [2017] found that the presence of stereotypes in gamified environments increased anxiety in women. Chang *et al*. [2019] revealed that women exposed to male avatars and sexist behaviors performed worse in mathematics. On the other hand, Santos *et al*. [2022a] observed an increase in aggression levels in participants exposed to stereotypes in gamified educational environments, especially among women. These studies underscore the importance of addressing and mitigating gender stereotypes in digital educational technologies to promote gender equality and enhance student performance (see **Table 1**).

Table 1. Comparisons among the studies

Study	Mediators	Gamification	Participants
[Christy and Fox, 2014]	Performance	Leaderboards	Women
[Albuquerque *et al*., 2017]	Anxiety and Performance	Avatars, Leaderboards, Colors	Women
[Chang *et al*., 2019]	Performance	Avatars and sexist behavior	Women
[Santos *et al*., 2022a]	Aggressiveness, Flow, Performance	Avatars, Leaderboards, Colors	Men and Women
[Nascimento et al. 2024]	Anxiety, Flow, Performance	Avatars, Leaderboards, Feedbacks	Men and Women

3 Experimental Design

Our study was developed from the question: “Does Gender Stereotype Boost and Threat Affect the flow experience, Self-Handicapping, and performance of Brazilian Students in gamified tutoring systems?” and aims to verify whether positively gender-stereotyped gamified educational environments affect the level of Self-handicapping and Flow of their users and whether this effect impacts performance and engagement in proposed activities. The participants in the experiment were high school and university students from public and private institutions in the state of Alagoas, located in the Northeast region of Brazil. The students were selected regardless of age, ethnicity, social class, and gender. The experiment involved tests administered on a gamified platform that contained stereotyped messages constructed from the study by Albuquerque *et al*. [2017]. The experiment followed a 2 x 3 factorial design, in which the first factor was the gender of the participants (men; women), and the second was the gamified environment to which they were allocated (Control Environment (Neutral); environment with Positive Male Stereotypes, and Positive Message for Men (P.M.S. – M.M.); environment with Positive Female Stereotypes, and Positive Message for Women (P.F.S. – M.W.)), see **Figure 1**. The dependent variables measured in the study were the level of Self-Handicapping, recorded before and after the main experimental task, the level of performance on an intellectual logic task, and the flow experience.

We consider the “stereotype threat condition” to be the situation in which the student is randomly assigned to an environment that does not correspond to their gender and the “stereotype boost condition” to be the situation in which the student is randomly assigned to an environment that corresponds to their gender expression. To achieve the objective of the study, we submitted the following hypotheses to empirical testing:

H1: There is no statistically significant difference between the measures of any of the participant groups for selfhandi-

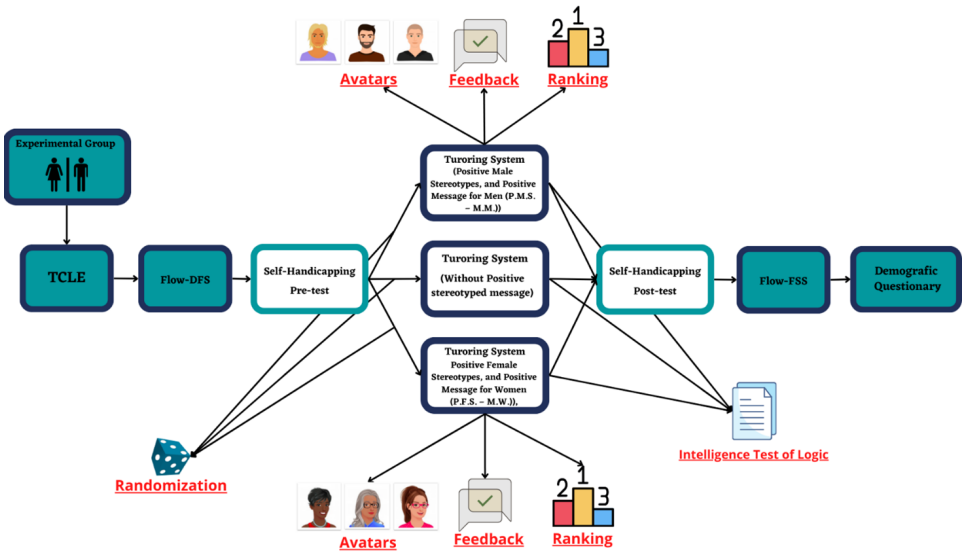

Figure 1. Design of experiment

capping. **H1.a** There is a statistically significant difference in self-handicapping when comparing the groups in stereotype threat condition and control condition; **H1.b** There is a statistically significant difference in self-handicapping when comparing the groups in stereotype boost condition and control condition; **H1.c** There is a statistically significant difference in self-handicapping when comparing the groups in the stereotype boost condition and stereotype threat condition;

H2: There is no statistically significant difference between the measures of any of the participant groups for Flow: **H2.a** There is a statistically significant difference in Flow when comparing the groups in stereotype threat condition and control condition; **H2.b** There is a statistically significant difference in Flow when comparing the groups in stereotype boost condition and control condition; **H2.c** There is a statistically significant difference in Flow when comparing the groups in stereotype threat condition and stereotype boost condition;

H3: There is no statistically significant difference between the measures of any of the participant groups for Performance. **H3.a** There is a statistically significant difference in Performance when comparing the groups in stereotype threat condition and control condition; **H3.b** There is a statistically significant difference in Performance when comparing the groups in stereotype boost condition and control condition; **H3.c** There is a statistically significant difference in Performance when comparing the groups in the stereotype threat condition and stereotype boost condition;

4 Method

4.1 Subjects

The sampling process was convenience-based and the subjects of the study were randomly allocated to the experimental conditions, according to **Table 1**.

Table 2. Subjects' Allocation: Gender X Platform

Contingency Table				
Gender	Control	AEF	AEM	Total
Woman	37	30	20	87
Man	20	21	19	60
Total	57	51	39	147

4.2 Subjects

In this section, we will describe the gamified environments and the instruments used to conduct the research.

4.2.1 Ethical Responsibilities

Following the provisions of Resolution No. 466 of December 12, 2012, which addresses the guidelines and norms regulating research involving human beings, always prioritizing human dignity, this research project was approved after submission to the Brazil Platform. Furthermore, considering the provisions in Item IV, "On the process of free and informed consent," of this resolution (466/2012), the following measures were taken by the authors:

Regarding underage students: High school participants were invited to take part in the research through mediated contact with one of their accompanying teachers. The teacher, who had a prior connection with the research group, was contacted via email and WhatsApp, and informed about the objectives and stages of the research. The teacher was instructed to invite students, who could only participate with the approval of their parents or guardians. No student was subjected against their will to this research, which was conducted with them remotely. All participants were presented with the Informed Consent Form (ICF), available on the research website, to be read and approved by the parents or guardians of minors. A simplified version was presented to students to facilitate Assent.

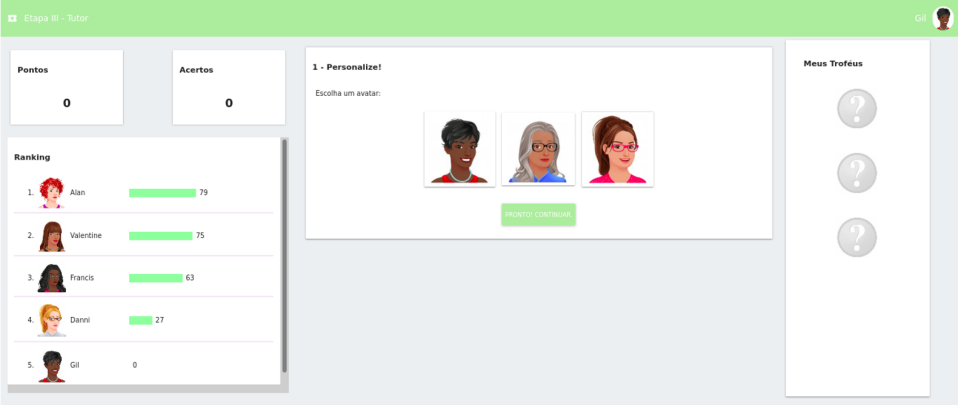

Figure 2. Female environment

Regarding undergraduate students: Undergraduate students were invited to participate in the research through visits made by the researcher to the locations where they carried out their academic activities, such as classrooms and research laboratories, as well as through emails sent to the ”periods” (equivalent to each class’s email) and messages sent to WhatsApp groups. They were also presented with the ICF, and the necessary ethical precautions were taken.

4.2.2 Gamified Educational Environments

The procedures were conducted in the following order:

1. Informed Consent Form;
2. Pre-Test for Self-Handicapping;
3. Flow Pre-Test
4. Logical Reasoning Questions;
5. Post-Test for Self-Handicapping;
6. Flow Post-Test;
7. Demographic Survey

Initially, upon accessing the experiment link, participants were redirected to the Informed Consent Form page. Upon questioning, they gave their consent for participation and publication of the results of their tests. They then proceeded to the Self-Handicapping tests, the gamified quiz, the Flow test, and a demographic survey. Participants had unlimited time to complete these activities. The Self-Handicapping measure was taken before (pre-test) and after (post-test) the gamified quiz, with 15 items in each version, respectively. After completing the pre-test of self-handicapping, students were randomly assigned to one of the three versions of the gamified environment: one with explicit positive male stereotypes, and positive feedback messages for men, another with explicit positive female stereotypes, and positive feedback messages for women, and another with no message at all. The gamified environment in which the quiz was conducted is based on the model developed by Albuquerque *et al*. [2017], but did not include implicit gender expression stereotypes in the colors and icons of the platform:

- Female environment **Figure 2**
- Male environment **Figure 3**
- Control environment **Figure 4**

The choice of colors was based on the study by Hallock [2003] regarding gender color preferences. In their study, with over 500 participants, men and women exhibited the same percentage of preference for green. Moreover, according to Thiel [2019], green is the color that represents harmony, balance, and comfort, and being used in institutions such as hospitals to calm patients. Positive stereotypes were employed in avatars, rankings (**Figure 2** and **Figure 3**), and after users resolved questions in the form of feedback: when answering a logic test question incorrectly, in male or female messaging platforms, the user received a positively stereotyped message about their gender – stereotyped by directly referring to participants who could identify as men or women – related to the study area of the activity being performed, see **Figure 5** and **Figure 6**. The message was crafted with the assistance of individuals of both genders, with prior knowledge of stereotype threat theory, to assess its appropriateness to the context and avoid promoting threat and lift. After completing the quiz, the user is redirected to the Self-handicapping post-test containing 15 questions, and upon completion, proceeds to the Flow test (FFS-2) containing 8 questions. After finishing the tests, the participant is invited to respond to a brief demographic survey.

4.3 Instruments

4.3.1 Self-handicapping

According to Mena [2019], the first Self-Handicapping scale was proposed by Martin [1998] and consisted of 27 items with a 4-point Likert scale. For this study, we used an adaptation of Mena [2019] version of the scale, presented in master’s thesis, which consisted of 19 items on a Likert scale (1 “Completely Disagree” to 4 “Completely Agree”). Our version used the 15 items with the highest added value for the pre-test - 10 items claimed for self-handicapping and five for active self-handicapping. For the post-test, versions opposite to the pre-test items were developed, adding 15 inverted items. The Likert scale ranged from 1 to 4, with 1 being “Completely Disagree” and five being “Completely Agree”. The variables we utilized were items: 3, 4, 9, 10, 11, 12, 15, 16, 17, and 25 for Claimed Self-handicapping, and 5, 6, 13, 14, and 27 for Active Self-handicapping [Mena, 2019].

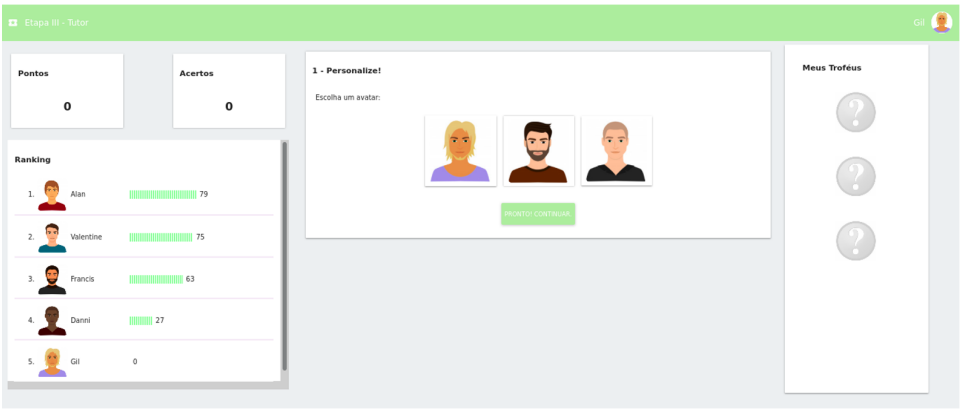

Figure 3. Male environment

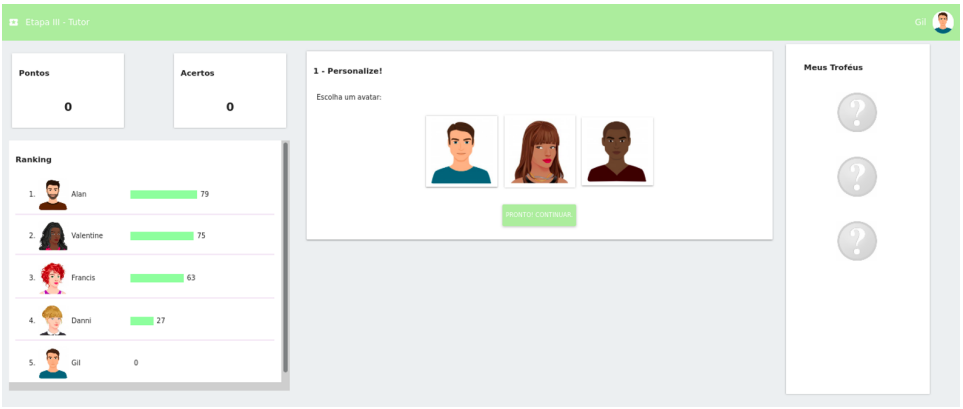

Figure 4. Control environment

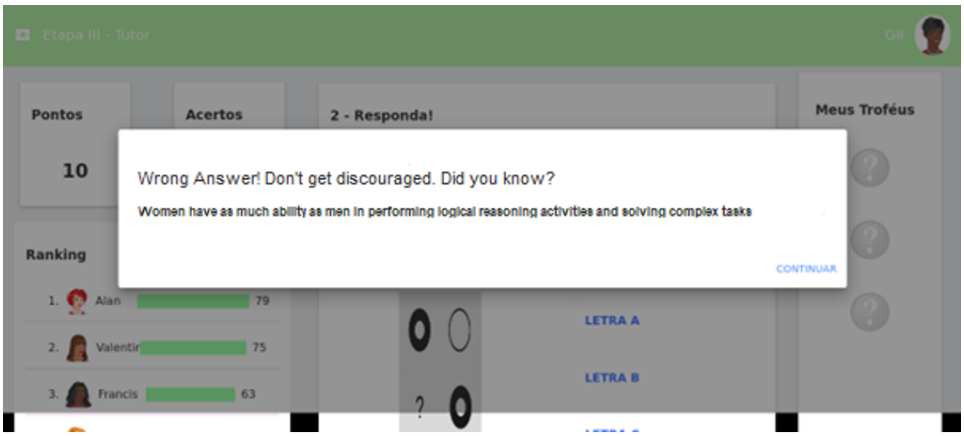

Figure 5. Women positive feedback

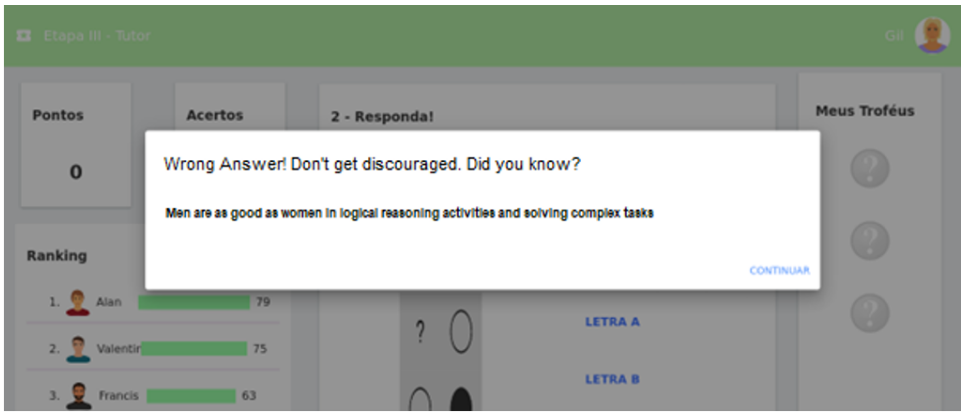

Figure 6. Men positive feedback

4.3.2 Flow

The applied Flow Test consisted of the short versions of the Brazilian Dispositional Flow Scale (DFS-2) and Flow State Scale (FFS-2), developed by **?**Bittencourt *et al*. [2024], which had nine each, aimed at measuring the level of engagement in the task. The Test was carried out in two stages, one before the pre-test of self-handicapping (DFS-2) and another at the end of the post-test of Self-Handicapping (FFS-2). The variables we used were items: 10, 20, 21, 4, 32, 24, 7, 17, and 36 for the Short version of the Dispositional Flow Scale (DFS-Short BR), and 4, 7, 10, 11, 23, 24, 26, 27, and 30 for the Short version of the Flow State Scale (FSS-Short BR) [Bittencourt *et al*., 2021, 2024].

4.3.3 Performance

The variables we used consisted of 20 visual logic questions, which involved identifying the item that correctly completed the figure. Participants started with a score of zero (0 points) and earned points as they answered the questions correctly. Initially, when selecting an avatar, they were awarded five points. Correct answers received 10 points, and incorrect answers did not score. When users reached 5 and 10 correct answers, they received a badge, which occurred when they finished the Quiz. An example of a question is presented in **Figure 7**.

The conduction of this experiment involves the following steps:

1. Answer and accept the terms of the study;
2. Answer the self-handicapping Test - Pre-self-handicapping Measure;
3. Answer the Dispositional Flow Scale (Short Version DFS-2)
4. At this moment, the system randomly generated either of the following versions of the platform for the participant: Control, Stereotypical Male Environment, or Stereotypical Female Environment;
5. Choice of avatars, according to the generated platform;
6. Answer the performance activity - Logical Test;
7. Answer the Self-handicapping Test - Post-Selfhandicapping Measure;
8. Answer the Flow State Scale (Short Version FSS-2).

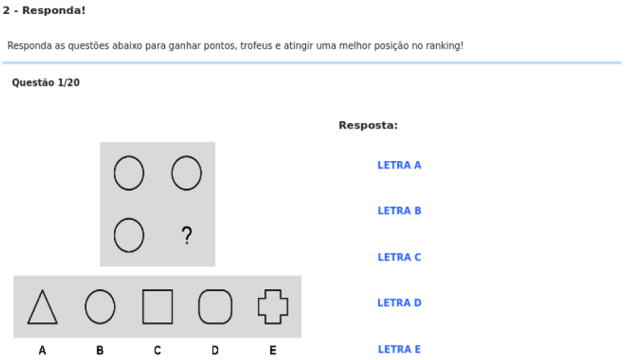

Figure 7. Logic question

5 Results

In this section, all the statistical analyses performed in this study will be presented. To ensure the validity of the statistical tests conducted, specific procedures were employed to verify whether the assumptions of each test were met. Firstly, a meticulous analysis was conducted to identify and remove potential outliers in the data. This process involved employing robust statistical criteria such as interquartile ranges and standard deviations, supplemented by visual assessment through scatter plots and boxplots. Furthermore, measures of skewness and kurtosis were calculated to assess the distribution of the data. These measures were compared against expected values for a normal distribution, and when necessary, transformations were considered to ensure that the assumptions of the statistical tests were satisfied. To mitigate the impact of outliers on the analysis, the technique of trimmed means was also employed, involving the removal of a percentage of extreme values and subsequent recalibration of measures of central tendency. These procedures were executed systematically and rigorously to ensure the validity and accurate interpretation of the statistical test results.

5.1 Self-handicapping

After controlling the linearity of covariance in the selfhandicapping pre-test (pre.shs), the ANCOVA test between the environment (env) (stMale, stFemale, neutral) and participant's gender (gender) (man, woman) were performed to determine statistically significant difference in the selfhandi-

Table 3. Results of Analysis of Covariance (ANCOVA) (env:gender) for participant's results considering their gender (gender) in self-handicapping in pre-test (pre.shs) and the self-handicapping post-test (shs) in stereotyped and non-stereotyped settings (env).

ANCOVA								
Variable	Effect	DFn	DFd	SSn	SSd	F	p	ges
shs	Pre.shs	1	140	29.360	81.039	50.721	<0.001	0.266
shs	env	2	140	0.524	81.039	0.453	0.637	0.006

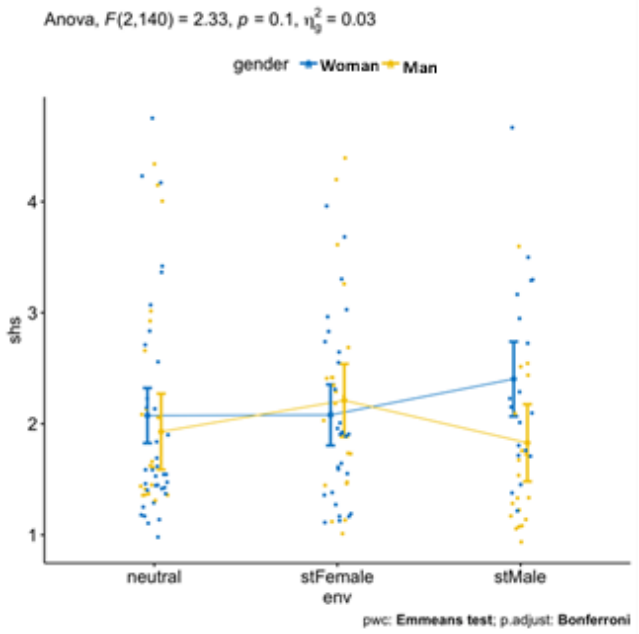

Figure 8. Analysis of Co-variance (ANCOVA) between participant's results in self-handicapping pre-test (pre.shs) and the self-handicapping post-test (shs) in stereotyped and non-stereotyped settings.

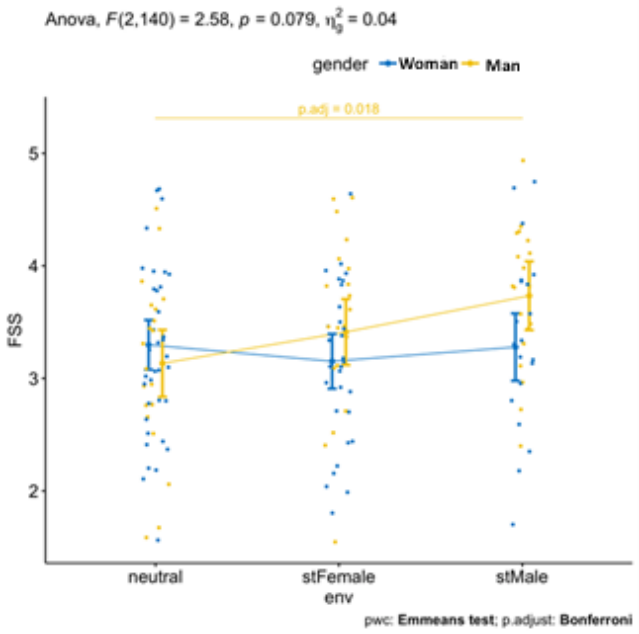

Figure 9. Analysis of Co-variance (ANCOVA) between participant's results in Dispositional Flow Scale (DFS) in pre-test and Flow State Scale (FSS) post-test in stereotyped and non-stereotyped settings.

capping post-test (shs).

Comparisons between conditions using Estimated Marginal Means (EMMs) were conducted to find statistically significant diferences adjusted p-values (see **Table 3**). The test indicated that there were statistically significant differences for the type of environment and participant gender (F(2,14)=2.33; **p=0.101**) (**Figure 8**). Thus, statistically, significant differences were found in the post-test measure of self-handicapping (shs) for participants for gender in stereotyped (stMale, stFemale) and non-stereotyped (neutral) conditions. This result rejects the null hypothesis H1.0, and confirms hypotheses **H1.a**, **H1.b**, and **H1.c**, regarding the existence of statistically significant differences in the levels of self-handicapping among participants when comparing the experimental groups, the experimental environment, and their gender.

5.2 Flow

After controlling the linearity of covariance in the predisposition to flow experience test (DFS), the ANCOVA test between the environment (env) (stMale, stFemale, neutral) and participant's gender (gender) (man, woman) was conducted to identify statistically significant difference in flow experience (FSS). Comparisons between conditions using the Estimated Marginal Means (EMMs) were computed to find statistically significant differences with adjusted p-values (see **Table 4**). The test indicated statistically significant differences for the type of environment and participant gender (F(2,14)=2.58; p=**0.079**). Thus, statistically significant differences were found in the post-test flow experience (FSS) for participants in stereotyped conditions (stMale, stFemale) and nonstereotyped conditions (neutral) regarding gender (**Figure 9**). This result reject the null hypothesis **H2.0**, and confirms hypotheses **H2.a**, **H2.b**, and **H2.c**, regarding the existence of statistically significant differences in the levels of Flow among participants when comparing the experimental groups, the experimental environment, and their gender

5.3 Performance

The ANOVA tests between the environment (env) (stMale, stFemale, neutral) and participant's gender (gender) (man, woman) was conducted to determine the statistically significant difference in the performance using activity points as metric (see **Table 5**). "activityPoints". The test did not indicate statistically significant effects for the performance variable on the interaction between the environment and gender (F(2.141)=2.83; p=**0.062**) (**Figure 10**). This result can not reject the null hypotheses **H3.0**, regarding the nonexistence of statistically significant differences in the levels of Performance among participants when comparing the experimental groups, the experimental environment, and their gender.

Table 4. Results of Analysis of Covariance (ANCOVA) (env:gender) for participant's results considering their gender (gender) in Dispositional Flow Scale (DFS) in pre-test and the Flow State Scale (FSS) post-test in stereotyped and non-stereotyped settings (env).

ANCOVA								
Variable	Effect	DFn	DFd	SSn	SSd	F	p	ges
FSS	DFS	1	140	14.664	63.415	32.373	**<0.001**	0.188
FSS	env	2	140	1.563	63.415	1.725	0.182	0.024
FSS	gender	1	140	0.875	63.415	1.933	0.167	0.014
FSS	env:gender	2	140	2.338	63.415	2.581	**0.079**	0.036

Table 5. Results of ANOVA between participant's performance (activity points) in stereotyped and non-stereotyped settings.

ANOVA								
Variable	Effect	DFn	DFd	SSn	SSd	F	p	ges
activityPoints	env	2	141	2329.108	325179.8	0.505	0.605	0.007
activityPoints	gender	1	141	1055.501	325179.8	0.458	0.500	0.003
activityPoints	env:gender	2	141	13090.166	325179.8	2.838	**0.062**	0.039

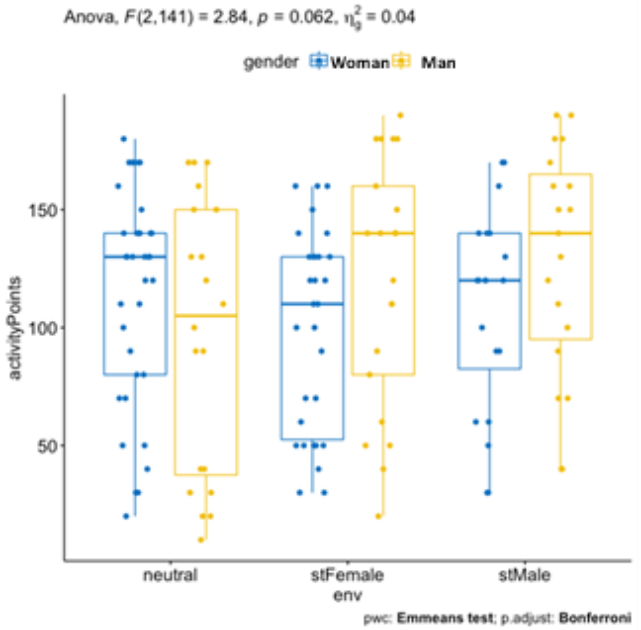

Figure 10. Analysis of Co-variance (ANCOVA) between participant's results in Dispositional Flow Scale (DFS) in pre-test and Flow State Scale (FSS) post-test in stereotyped and non-stereotyped settings.

6 Discussion

Hypothesis **H1.0** was rejected due to the confirmation of hypotheses **H1.a**, **H1.b**, and **H1.c**, which were related to the measurement of self-handicapping for the different participant groups in the study. Statistically significant differences were found when comparing the measurements of the three groups. It is interesting to note that among the control and stereotype threat conditions, both men and women exhibited a higher level of self-handicapping when exposed to the stereotype threat condition, but the measurement for women was higher. This result is consistent with findings in the literature, such as those from Christy and Fox [2014], Pennington *et al*. [2016], Albuquerque *et al*. [2017], Chang *et al*. [2019] and Santos *et al*. [2022a], which point to a negative effect of stereotype threat on psychological mediators. When comparing the stereotype boost situations, it was observed that self-handicapping decreased in men, as well as in women, when compared to the control group. This result indicates, as proposed by Shih *et al*. [2002], Smith and Johnson [2006], and Swift *et al*. [2013], that stereotypes can also have positive effects when applied correctly, by valuing the social identity of the group to which the individual belongs and consequently reducing effects that may impair learning.

Hypothesis **H2.0** was rejected based on the confirmation of hypotheses **H2.a**, **H2.b**, and **H2.c**, which addressed the Flow measurement for the different participant groups. Statistically significant differences were found in all cases. When comparing Flow measurements for stereotype threat conditions, it was found that groups under threat developed a greater measure of Flow when compared to the neutral environment. Men under threat exhibited a higher level of flow, as did women, when compared to their peers in the neutral environment. These results reflect part of the multiple stereotype threat model theory, in which individuals, even under threat, may not experience performance loss or negative influences on cognition, and may even feel motivated by being in that condition [Shapiro and Neuberg, 2007]. When comparing the stereotype boost situations, an interesting phenomenon was observed. Women in the boost condition showed a lower measure of flow compared to their peers when exposed to the threatening environment. On the other

hand, men had an even higher measure of flow than their peers in the threatening environment. These results respectively indicate the multiple stereotype threat model theory by Shapiro and Neuberg [2007] and the influence of positive stereotypes [Shih *et al*., 2002; Smith and Johnson, 2006; Swift *et al*., 2013]. Hypothesis **H3.0** was not rejected; consequently, hypotheses **H3.a**, **H3.b**, and **H3.c** were rejected as statistically significant differences were not found regarding participants' performance when comparing the results of the gamified quiz through the environment and the stereotype boost or threat condition. This could have two explanations. The first would be the difficulty level of the quiz, which may not have been challenging enough or was at a complexity level that students could handle. These results contradict findings from [Christy and Fox, 2014; Pennington *et al*., 2016; Chang *et al*., 2019] which report impacts on performance following participants' exposure to stereotype threat conditions.

7 Limitations

Firstly, due to the methodology employed for data collection, comprehensive monitoring of all test applications wasn't feasible. Consequently, it remains unknown whether participants engaging remotely employed strategies to cheat on the gamified logic test, potentially influencing performance results. Secondly, the concepts of gender and racial variability utilized in this study may be constrained in their conception, potentially reinforcing stereotypes. This limitation arises from the prevalence of a heteronormative social organizational logic, which shapes perceptions of individual identity. Consequently, both avatar appearance and messages may implicitly exhibit cisgender bias. Moreover, the study's generalizability is challenged by the diversity of participants, encompassing high school and undergraduate students. Furthermore, the convenience sampling method, restricted to a single state in a specific region of Brazil, limits the broader applicability of the results. Lastly, the extensive length of the self-handicapping measurement questionnaires may have induced discomfort and fatigue among participants during the completion process, potentially influencing responses.

8 Future works

For future studies, the authors recommend using platforms better adapted to the mobile context, as the web application may have been insufficient during the experiment, affecting the visibility of items and consequently the user experience in the gamified educational environment. We also recommend conducting more studies on the impact of stereotypes in educational technologies, as well as other social categories that may be inherently related to learning problems. As a direction for future research, a more in-depth exploration of the specific characteristics of gender stereotypes, utilizing both quantitative and qualitative methodologies, is recommended. Additionally, investigating pedagogical strategies that can modulate these effects to optimize the educational environment and promote gender equity is crucial. A more refined understanding of these dynamics will contribute to the development of more inclusive and effective pedagogical practices in gamified environments, representing a significant advancement in the field of computer science in education. One possibility within this perspective is to develop machine learning and artificial intelligence models that, upon detecting specific user characteristics, can dynamically adapt the educational environment in real-time, aiming to mitigate predominant characteristics that may exclude unidentified groups.

9 Conclusion

complexity of the presented tasks. These findings have significant implications for both academia and industry. Our results contribute to advancing knowledge on how gender stereotypes affect gamified educational environments, expanding the field of study in educational psychology and educational technology. Moreover, they underscore the importance of considering gender dynamics in the development of more inclusive and effective teaching methods, as well as gender sensitivity in the development of curricula and educational materials. Furthermore, our results suggest that designers of gamified educational platforms should carefully consider how gender stereotypes may influence user experience and academic performance, aiming for the development of more inclusive platforms. Additionally, they can inform teacher training programs, empowering educators to recognize and address gender stereotypes in their teaching practices, and marketing and engagement strategies, enabling companies to develop products that better meet the needs of different demographic groups.

Declarations

Acknowledgements

We would like to thank our colleagues, and members of NEES, Center of Excellence in Social Technologies, for their support and suggestions during the conduct and construction of this work.

Authors' Contributions

JVLBN contributed to the conception of this study and performed the experiments. JJS, and IIB contributed with the design of methodology. GCC contributed with the Application of statistical. JVLBN is the main contributor and writer of this manuscript. All authors read and approved the final manuscript. .

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the Harvard Dataverse repository, https://doi.org/10.7910/DVN/D9TTSL.

References

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

Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., and Doms, M. E. (2011). Women in stem: A gender gap to innovation. Technical Report 04-11, Economics and Statistics Administration Issue Brief. Available at:https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1964782.

Bittencourt, I. I., Freires, L., Lu, Y., Challco, G. C., Fernandes, S., Coelho, J., Costa, J., Pian, Y., Marinho, A., and Isotani, S. (2021). Validation and psychometric properties of the brazilian-portuguese dispositional flow scale 2 (dfs-br). *PloS one*, 16(7):e0253044. DOI: 10.1371/journal.pone.0253044.

Bittencourt, I. I., Freires, L., Lu, Y., Challco, G. C., Fernandes, S., Coelho, J., *et al*. (2024). Psychometric properties of the brazilian-portuguese flow state scale short (fss-br-s). *PloS ONE*, 19(2):e0286639. DOI: 10.1371/journal.pone.0286639.

Chang, F., Luo, M., Walton, G., Aguilar, L., and Bailenson, J. (2019). Stereotype threat in virtual learning environments: Effects of avatar gender and sexist behavior on women's math learning outcomes. *Cyberpsychology, Behavior, and Social Networking*, 22:634–640. DOI: 10.1089/cyber.2019.0106.

Christy, K. R. and 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. DOI: 10.1016/j.compedu.2014.05.005.

Costa, A. and Marchiori, P. (2015). Gamificação, elementos de jogos e estratégia: uma matriz de referência. *InCID: Revista de Ciência da Informação e Documentação*, 6:44. DOI: 10.11606/issn.2178-2075.v6i2p44-65.

Csikszentmihalyi, M. (2020). *Flow (Edição revista e atualizada): A psicologia do alto desempenho e da felicidade*. Objetiva. Book.

Hallock, J. (2003). Colour assignment - preferences and associations. Available at:http://www.joehallock.com/edu/COM498/credits.html.

Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. Pfeiffer. Book.

Kim, J. and Castelli, D. (2021). Effects of gamification on behavioral change in education: A meta-analysis. *International Journal of Environmental Research and Public Health*, 18:3550. DOI: 10.3390/ijerph18073550.

Martin, A. J. (1998). *Self-Handicapping and defensive pessimism: Predictors and consequences from a self-worth motivation perspective*. PhD thesis, University of Western Sydney, Macarthur. Available at:https://researchdirect.westernsydney.edu.au/islandora/object/uws:587/.

Mena, R. O. S. (2019). Estudo das relações entre as estratégias de self-handicapping, autoestima, autoeficácia e o rendimento académico: Um estudo com alunos do ensino superior. Master's thesis, Universidade da Madeira, Funchal. Available at:http://hdl.handle.net/10400.13/2700.

Moreira, G. L. R. and F., K. G. (2012). *Level 4: The gamification toolkit: Game element*. Wharton Digital Press. DOI: 10.9783/9781613631041-005.

Moreira, G. L. R. and F., K. G. (2018). Elementos de gamificação no aplicativo duolingo. *Disciplinarum Scientia*, 19(2):205–214. Available at:https://periodicos.ufn.edu.br/index.php/disciplinarumCH/article/view/2923.

Myers, D. G. (2014). *Preconceito*, pages 246–277. AMGH, Porto Alegre, RS. Book.

Pennington, C. R., Heim, D., Levy, A. R., and Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PLOS ONE*, 11(1):1–25. DOI: 10.1371/journal.pone.0146487.

Piatek-Jimenez, K., Cribbs, J., and Gill, N. (2018). College students' perceptions of gender stereotypes: Making connections to the underrepresentation of women in stem fields. *International Journal of Science Education*, 40(12):1432–1454. DOI: 10.1080/09500693.2018.1482027.

Santos, J., Benevides, K., Andrade, E., Nascimento, J., Silva, K., Pinto, I. I. B. S., *et al*. (2022a). Does gender stereotype threat affects the levels of aggressiveness, learning and flow in gamified learning environments?: An experimental study. *Educ Inf Technol*, pages –. DOI: 10.1007/s10639-022-11220-3.

Santos, J., Benevides, K., Andrade, E., Nascimento, J., Silva, K., Pinto, I. I. B. S., *et al*. (2022b). Tecnologias educacionais estereotipadas: Um desafio a ser enfrentado. *Revista Brasileira de Informática na Educação*. DOI: 10.5753/rbie.2022.2293.

Shapiro, J. R. and Neuberg, S. L. (2007). From stereotype threat to stereotype threats: Implications of a multithreat framework for causes, moderators, mediators, consequences, and interventions. *Personality and Social Psychology Review*, 11(2):107–130. DOI: 10.1177/1088868306294790.

Shih, M., Ambady, N., Richeson, J. A., Fujita, K., and Gray, H. M. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology*, 83(3):638–647. DOI: 10.1037/0022-3514.83.3.638.

Shih, M., Pittinsky, T. L., Trahan, A. G., and Legault, L. (2012). Stereotype boost: Positive outcomes from the activation of positive stereotypes. In *Stereotype threat: Theory, process, and application*, pages 141–156. Oxford University Press. Available at:https://psycnet.apa.org/record/2011-27639-008.

Smith, J. L. and Johnson, C. S. (2006). A stereotype boost or choking under pressure? positive gender stereotypes and men who are low in domain identification. *Basic and Applied Social Psychology*, 28(1):51–63. DOI: $10.1207/s15324834basp2801_5$.

Snyder, C., Lopez, S., and Lopez, S. (2009). Mindfulness,

flow e espiritualidade: em busca das melhores experiências. In *Psicologia positiva: uma abordagem científica e prática das qualidades humanas*, pages 222–238. Artmed, Porto Alegre. Book.

Steele, C. M. and Aronson, J. (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of Personality and Social Psychology*, 69(5):797–811. DOI: 10.1037/0022-3514.69.5.797.

Swift, H. J., Abrams, D., and Marques, S. (2013). Threat or boost? social comparison affects older people's performance differently depending on task domain. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, 68(1):23–30. DOI: 10.1093/geronb/gbs044.

Thiel, C. (2019). *A Psicologia das Cores no Marketing*. Ebook. Book.

Török, L., Szabó, Z. P., and Tóth, L. (2018). A critical review of the literature on academic self-handicapping: theory, manifestations, prevention and measurement. *Social Psychology of Education*, 21:1175–1202. DOI: 10.1007/s11218-0189460-z.

Zanatto, R. (2007). Perfil motivacional de alunos de arquitetura : um estudo exploratório. Master's thesis, Universidade Estadual de Londrina, Londrina. Available at:https://bdtd.ibict.br/vufind/Record/UEL_b74718cb77ef4de95666b54c7d7e6beb.