Education and Information Technologies (2023) 28:1637–1662
https://doi.org/10.1007/s10639-022-11220-3

Does gender stereotype threat affects the levels of aggressiveness, learning and flow in gamified learning environments?: An experimental study

Jário Santos[1] · Ester Andrade[2] · Kamila Benevides[2] · Kelly Silva[2] · João Nascimento[3] · Ig Bittencourt[2,3] · Marcos Pereira[4] · Sheyla Fernandes[3] · Seiji Isotani[1]

Received: 28 January 2022 / Accepted: 3 July 2022 / Published online: 1 August 2022

Abstract

Studies in the literature reported several positive benefits provided by the use of technology in online education, especially in the gamified tutoring system. However, despite the benefits of intelligent tutoring systems, recent studies indicate the presence of a gender gap not considered in the construction of the attributes present in the gamified tutoring system. To investigate this impact by observing users' behavioral changes in gamified online educational environments, the present study aims to investigate the effects of the stereotype threats using a quantitative experiment with a Factorial Design in three gamified environments (stereotypical male version, stereotypical female version and control environment). Was conducted an experiment with 150 individuals (high school and undergraduate students) without considering age, ethnicity, or social class. The results show that the participants allocated to the male learning environment present an increase in aggressiveness level. Furthermore, the results also show the stereotypical male and female learning environments increased the participants' performance level. Another finding was that the threatening condition provided a significant increase in the participants' flow level among males subjected to a threatening condition, which did not manifest in the case of females. In addition, this study also observed the effect of the stereotype threat on men and women in the threatening condition by division in the 34-year age group, resulting in a significant increase in the level of flow among men. This study showed previous results show that the gamified environment influences psychological variables as aggressiveness, intellectual performance, and flow level, they raise questions about the direction of these changes and the impact they may have on users' usability and performance in these systems.

Keywords Stereotyped educational environments · Stereotype threat · Intelligent tutoring systems · Gender stereotype threat · Aggressiveness

Extended author information available on the last page of the article

1 Introduction

Studies in the literature reported several positive benefits provided by the use of technology in online education (Cheung & Slavin, 2013) (Hwang, 2003) (Kizilcec et al., 2017) (Kappen et al., 2019) (Friedman et al., 2008), such as decentralization with new directions and teaching strategies (Francescucci et al., 2020), mainly with a focus on promoting learning (Bailey & Lee, n.d.Kantharia, 2020). Studies have several positive characteristics such as (i) personalization of teaching and learning (Kizilcec & Lee, 2020), favoring and assisting the student in building knowledge efficiently; (ii) promotion of students' development and responsibility (Aguilar et al., 2013); and (iii) provision of education anywhere, anytime, and for anyone (Bittencourt et al., 2008), and it is mediated by educational systems.

Intelligent tutoring systems present several pedagogical benefits for students. The primary advantage offered by intelligent tutoring systems is the personalized monitoring of learning (Anderson et al., 1985), which generates a significant improvement in the students' learning performance. These positive benefits are enhanced when gamification is implemented in intelligent tutoring systems, which adds motivational benefits to the learning process (De Sousa Borges et al., 2014). In this scenario, several gamification elements (trophies, rankings, points, and levels) (Dicheva et al., 2015) are used for the development of an Intelligent Tutor System, making it even more efficient for students' learning and motivational aspects.

However, despite the benefits of intelligent tutoring systems, recent studies indicate the presence of a gender gap not considered in the construction of the attributes present in the gamified tutoring system (Albuquerque et al., 2017; Kaye et al., 2018; Liu et al., 2021; Cross et al., 2022). As a result, instead of increasing students' learning performance and motivational aspects mediated by attributes in the educational system, the effect is the opposite, attributing advantages and disadvantages to specific genders. This problem is known as the Gender Stereotype Threat.

The impacts of the stereotype threat are related to several mediators (Pennington et al., 2016): (i) affective mechanisms; (ii) cognitive mechanisms; and (iii) motivational mechanisms. In affective mediators, studies classify and contextualize the stereotype threat linked to anxiety, performance expectations, individual characteristics, or even fear of being evaluated. In cognitive mechanisms there are works demonstrating some mediators such as working memory, cognitive load, and suppression of thought, is related to the low performance of women in mathematical activities, they feel more threatened when they know that they are being subjected to tests according to your gender. Some motivational mechanisms, such as effort/motivation, suggest that the stereotype threat, in addition to impacting the performance of some students, acts with the opposite effect: discouraging or generating a high load of effort for the development of an activity that could be considered simple.

There are many cognitive mediators and when the subject is under threat, the negative effects can even cause a decrease in his performance (Pennington et al., 2016). Studies observe the relationship of these cognitive mediators under the

effects of anxiety (Albuquerque et al., 2017). However, it is important to point out that the stereotype threat in certain situations, such as direct exposure to the stereotype before an academic activity, can result in increased levels of aggression, and this at high levels acts at the expense of learning and the ability to concentration (Inzlicht & Kang, 2010). In addition, based on Albuquerque et al. (2017) and Pennington et al. (2016), the main research question about this study "What is the impact that aggressiveness, mediated by the Stereotype Threat, can have on the e-learning processes of users of a Gamified Platform?". This investigation is important, since, as seen in Albuquerque et al. (2017), through the effects of the threat of stereotypes, a Gamified Educational Platform can act against the purpose for which it was created, causing a decrease in the levels of motivation, engagement and consequently of learning. This would occur due to the presence of stereotyped elements for specific genders, such as colors, rankings and badges.

In view of the findings, of Inzlicht and Kang (2010) and Pennington et al. (2016), and a systematic literature review on the topic, in which no article was found that mentioned or proposed to mediator analysis Aggressiveness in a Gamified Educational Platform, it is necessary to investigate this gap in the literature, since it is possible that the elements of a gamified platform can act against the purposes for which it was created. Furthermore, the effects provoked by individuals in situations of stereotype threat under the mediation of aggression can enter boredom states, and as a consequence, be related to a drop in the flow experience (Clarke and Haworth, 1994).

The gender stereotype threat needs to be further investigated due to its several negative effects, such as low learning performance, which is mediated by factors that also affect engagement and motivational aspects (Pennington et al., 2016). Therefore, the purpose of this paper is to identify the relationships between the users' level of aggressiveness, intellectual performance, and the flow experience in a gamified educational environment. To achieve this objective, we used the aggressiveness questionnaire (Buss & Perry, 1992) to measure this mediator, the intellectual performance was measured using an instrument specially designed for the experiment (logic test), consisting of 20 items and the flow measure was obtained using a questionnaire containing 26 items. Quantitative, descriptive, and inferential analyzes were used to find and discuss the results.

This article is structured as follows: Section 2 provides the theoretical background and the related works. In Section 3, we present the proposal and the tool used in this study. In Section 4, we depict the results obtained in the conducted experiment. In Section 5, we develop a discussion about the results previously reported. Finally, in Section 6, we provide our concluding remarks.

2 Background

In this section is presented main concepts about the proposed work. In addition, we are going to depict related works to this study, more precisely related to gamified educational environments, stereotype threat in virtual environments.

2.1 Stereotype threat

The stereotype threat is characterized as a form of threat to the social identity, and it is triggered by a situation in which a negative stereotype is incurred by a social group that is being devalued by, or in relation to, another (Pennington et al., 2016). The reading of these facts and their explanation is related to the cultural context, and to an intrinsic concept about an external behavior, which would be linked to the formation and maintenance of stereotypes since values and norms are transmitted by generations and justified in the preservation of cultural traditions (Guerra, 2002).

This situation of identity conflict represents a threat to personal integrity, in which individuals are inhibited from performing optimally by the awareness that a possible failure can cause a negative judgment, which can generate a drop in cognitive performance and the expectation of success (Steele, 1988). Studies in the literature conceptualize two dimensions of the stereotype threat: one related to the targets of the threat (self or his/her group) and the other to the sources of the threat (self, internal group, or external group). Figure 1 describes the intersection of these dimensions, demonstrated in the research (Shapiro & Neuberg, 2007). According to it, situations can be incited by the possibility that the behavior and/or individual differences interfere actively in the conceptualization about the individual or the group, increasing the susceptibility of negative stereotype. As a consequence, individuals in this situation may develop feelings such as anxiety and apprehension.

Fig. 1 Source of the threat x Target of the threat. Shapiro and Neuberg (2007)

In the systematic literature review of [10], a theory regarding members of different stigmatized groups were reported. According to this theory, individuals can report different ways of experiencing stereotype threat and these distinct experiences can be mediated by different processes. Although these different forms of stereotype threat may have the same negative effects on performance, they are likely to be stimulated by different mechanisms, so it is necessary to evaluate whether the same mechanisms are the factors responsible for the effect of the threat on different groups.

In the systematic literature review of Pennington et al. (2016), a theory regarding members of different stigmatized groups were reported. According to this theory, individuals can report different ways of experiencing stereotype threat and these distinct experiences can be mediated by different processes. Although these different forms of stereotype threat may have the same negative effects on performance, they are likely to be stimulated by different mechanisms, so it is necessary to evaluate whether the same mechanisms are the factors responsible for the effect of the threat on different groups.

2.2 Mediator: Aggressiveness

The article (Inzlicht et al., 2011) suggests that the stereotype threat has a long-term influence, which results in behavioral difficulties in individuals, even after leaving the environments where they feel the threat. Based on it, this article brings the theory of stereotype threat spillover, which comprises a broader understanding of how stereotype threat spillover operates. According to this perspective, the threats reach not only the domains where people suffered this type of judgment, but also reach domains that are usually considered stereotype-free. The theory of Inzlicht et al. (2011) is based on the premise that time spent in an intimidating environment causes emotional and cognitive distress, and these residual effects influence behavior.

Presenting the same theoretical perspective, which stereotype and social identity threat result in a series of physiological, emotional, cognitive, and behavioral consequences, another research (Inzlicht and Kang, 2010) also exposes the stereotype threat as a source of stress. When becoming more prone to failure, the individual has physiological responses linked to stress because of the created expectation and negative thoughts that consume their cognitive abilities (Inzlicht & Kang, 2010). Therefore, Inzlicht et al. (2011) concludes that dealing with stereotype stress and the threat to social identity influences the emotional field and, consequently, decreases performance in stressful situations.

The research (Inzlicht and Kang, 2010) also suggests that these residual effects, pointed out by Inzlicht et al. (2011), generate aggressive behaviors, and stereotype threat acts as a stressor in any situation that the individual has to be involved. The results of the study point out that the more people are aware that they are being evaluated, the more aggressive they become towards the person who is evaluating. Moreover, the study also shows that one of the causes of this phenomenon is related to the negative feedback received. Finally, the study points out that the lack

of capacity to control the impulses generated by anger is a feeling generated when facing the stereotype threat.

Some problems mentioned above, caused by the stereotype threat, are related to the increase of the aggressiveness level, showing the need to understand the factors and the basic concepts of aggressiveness in order to mitigate or stop this threat. The study conducted in Anderson and Bushman (2002) brings the main theories related to aggression, highlighting the Cognitive Neo-association Theory. According to this theory, a sequence of behaviors such as frustration, provocations, loud noises, uncomfortable temperatures, and unpleasant odors can produce negative effects. These effects are produced by unpleasant experiences, and automatically trigger thoughts, memories, physiological responses, motor reactions, and emotional reactions. However, these same causes provoke the activation of priming (Anderson & Bushman, 2002) and spread to other types of aggression.

Other studies approach the aggressiveness related to the instructors' communication in the educational environment, investigating how the evaluative messages or feedbacks are expressed. The research (Myers & Rocca, 2000) identifies the instructor's form of verbal aggressiveness. This aggressiveness is identified by the students as responsible for provoking several distinct impressions that can affect their performance and engagement. This indicates that instructors need to be aware of their behaviors and their way of communicating because their pedagogical skills can be improved in the educational environment, making the experience more positive.

In this sense, the stereotype threats are configured as a set of artificial dangers that can generate hostile environments. In the educational context, some situations may arise, and certain social groups, with tendencies to suffer stereotype threat, can awaken the consciousness of the risk of failure, of being harmed, or of having lower performance. These situations can cause an increase in the level of aggressiveness due to the presence of judgments and discriminatory behaviors.

2.3 Stereotyped gamified educational environments

Gamification is the application of game elements (e.g., competition, collaboration, rewards, scores, interaction, levels, phases) in non-game contexts (Deterding et al., 2011; Sheldon, 2012). This approach has been extensively used in education, aiming to increase students' engagement, change students' behavior, and improve students' learning outcomes (De Sousa Borges et al., 2014).

Although gamification is an agent of pedagogical change, the article (Orji et al., 2018) shows a different perspective concerning the effects of gamified systems. These environments usually adopt a "one-size-fits-all" approach and disregard the individual peculiarities during the learning process. Therefore, in this context, users are more likely to suffer stereotype threats.

Some studies conducted in gamified environments point out the performance difference concerning gender (Yücel & Rizvanoglu, 2019), reporting women to feel uncomfortable in such environments. A possible reason for this result is the low use of computers by women (Colley & Comber, 2003) and the decrease of women in STEM fields. In the study (Ortner & Sieverding, 2008), they used an emotion

software (Facial Action Coding System) to measure the feelings of girls while performing activities in STEM fields. The study revealed that texts with male stereotypes arouse negative emotions in females.

Therefore, even though gamification offers a variety of benefits (e.g., motivation and engagement during the learning process), if the platform is not an inclusive learning environment, it can have a detrimental effect on students' performance. This negative effect can cause a decrease in productivity, an increase in dropout numbers, and other consequences generated by stereotype threats.

2.4 Flow experience

The flow theory of Csikszentmihalyi (2020) has been applied in different contexts, be they education, culture or sports (Csikszentmihalyi, 2020; Snyder et al., 2009). The flow state is characterized as a state of constant fruition, in which the individual reaches a high concentration and has high intrinsic motivation. For the task to be considered as a Flow promoter, it needs to provide the individual with at least one of the following eight components (Csikszentmihalyi, 2020): 1. "balance between challenge and skill", 2. "fusion between action and awareness", 3. "clear goals", 4. "feedback", 5. "total concentration on the task of the moment", 6. " paradox of control", 7. "loss of self-awareness", 8. "transformation of time". The systematic review employed by dos Santos et al. (2018) indicates that there is positive evidence regarding the application of flow theory in the fields of computing and education, especially the version by Csikszenmihaly (dos Santos et al., 2018). In these scenarios, theory is being increasingly used and presents itself as an essential factor to generate motivation and promote increased learning and the satisfaction of users of educational systems (computer-based learning) (dos Santos et al., 2018). Thus, it is necessary to take into account the conditions to reach the state of flow when promoting an activity in an educational environment. For this, there are several strategies, including the gamification of these virtual spaces.

2.5 Related works and proposed work

Although the continuous search for alternative approaches by teachers to increase students' motivation and engagement (Sailer and Homner, 2020), as perceptions on the effectiveness of active learning strategies (Daouk et al., 2016), the traditional education system still shows the ineffectiveness of the passive learning method (Riley & Ward, 2017), which is widely used (Dicheva et al., 2015). Therefore, gamification emerges as an alternative to traditional teaching methodologies, presenting game elements in educational platforms, from its context to the design strategies and environment modeling. In gamification, components such as ranking, levels, and rewards are used and can be applied in various areas such as marketing, health, politics, education, among others (Dicheva et al., 2015).

As an example of the efficiency of introducing gamification in teaching-learning environments, we have the study of Gómez Contreras (2020). This study aimed to contribute to the increase of the application of gamification, considering its

theoretical foundations, in educational learning systems. The study was conducted in a distance course in Public Accounting at the Military University of New Granada. The work revealed that the environment implemented with gamification promoted more autonomy, social interaction, motivation and commitment among students.

Another example of work involving gamified educational platforms is Pedro and Isotani (2016). In this study, a gamified educational environment is presented, called E-Game, which aims to reduce inappropriate behaviors called gaming the system (Cetintas et al., 2009), a behavior that consists of attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning the material. This study conducted experiments with elementary school students, and two platforms were used: a gamified system and a non-gamified system. At the end of the study, they concluded that the gamified platform allowed the decrease of ”gaming the system” behaviors. Furthermore, behavioral and motivational differences between genders were detected since boys showed a higher level of engagement in the gamified environment when compared to the girls' level of engagement, and this occurred because of the presence of a feeling of incapacity in girls inserted into these environments. The article (Pedro & Isotani, 2016) is highly relevant because it allows us to think and consider the presence of the Stereotype Threat in gamified educational systems because of this disparity detected between the genres.

In the Threat context, a systematic literature review was conducted to investigate research involving stereotype threat (Pennington et al., 2016). This review identified the main psychological mediators related to this type of threat and reinforces the concept that multiple factors can affect academic performance, either in tasks of high or low complexity. However, the study emphasizes that stereotype threat justifies many results below the expected in certain social groups, depending on the level of achievement in certain social groups, depending on the position they occupy and the way in which the activity is constructed. In another study (Pennington et al., 2016), concerning the investigation of the impact of gender-related stereotypes on females in online communities of electronic games, it was pointed out that the subjects do not experience the stereotype threat in the same way, but in different ways depending on their level of identification and involvement with the group to which they belong.

In Kaye and Pennington (2016) was conducted a study that aims to analyze the impact of stereotype threat on the performance of females playing online games. The results pointed out that females who were under stereotype threat underperformed on the gaming task relative to males. Moreover, this study suggests that the performance differences could be eliminated when females identify with an alternative positive social identity. In Inzlicht and Kang (2010) was investigated the effects of stereotype threat to social identity. The study indicates evidence pointing to an increase in aggressiveness, unhealthy food consumption, risky decision-making, and attention control when subjects are faced with a stereotype-threatening situation. Therefore, based on the results reported in the studies that addressed stereotype threats, we can relate the low performance of the girls identified in the educational platform of the study (Pedro & Isotani, 2016) with the possible elements of threat that could be present in it.

In Laureano-Cruces et al. (2016), it is addressed the topic of modeling intelligent learning systems. This study considers color psychology, learning styles, and the role of emotions during the learning process to develop an intelligent learning system based on affective computing theories to provide a better experience for students. Color psychology is an area that deals with the importance of colors in the existing components of educational environments, and this psychology can awaken motivational responses in users, since each color has its particularity and different forms of perception and communication. For example, the color red is related to aggressiveness, hate and also serves to send messages of danger and attention to learning systems. In another study, Vakili et al. (2019) was investigated the effect of the color red in the classroom environment. In this study, it was pointed out that students exposed to classroom environments with red walls had an increased level of aggressiveness compared to students who were in white environments, for example.

In Lopes et al. (2019) was proposed a model for an automated gamified system based on the user profile and its interaction with the environment, aiming to solve the presence of possible stereotype threats. Considering that users access the system for different motivations and that these can change over time, in this proposed system model, a personality test would occur at the beginning of its use, and during the usage time, users could make changes based on their personal preferences and would receive tips and suggestions for customization. This proposal related to the creation of a model for an automated gamified system is interesting because it aims to improve users' motivation and engagement based on their profile and interaction with the platform.

The objective of this study is to identify the relationships between the users' level of aggressiveness, intellectual performance, and the flow experience in a gamified educational environment, considering the stereotypical or non-stereotypical nature of the environment and threats to users' identity. The first factor was the gender of the participants (male; female), and the second was the gamified environment where the students were allocated (control; stereotyped male (aem); stereotyped female (aef)). The dependent variables measured in the research were the level of aggressiveness (recorded before and after the main experimental task), the level of performance in an intellectual task, and the flow experience, according to Fig. 2.

To achieve the objective of the study, we submitted the following hypotheses to empirical testing:

H1: Stereotyped environments provide distinct levels of experience for users, with reflections on aggressiveness, performance, and flow measures. The formalization of this hypothesis can be expressed in the following terms:

$H_{1.a}$ Post Aggressiveness Level(control $\neq$ aem = aef): The level of Aggressiveness post-interaction is different in the control group when compared to aem and aef. The level of aggressiveness between aem and aef is the same;
$H_{1.b}$ Performance(control $<$ aem = aef): The performance obtained by solving activity is lower in the control group when compared to aem and aef. The performance between aem and aef is equal;

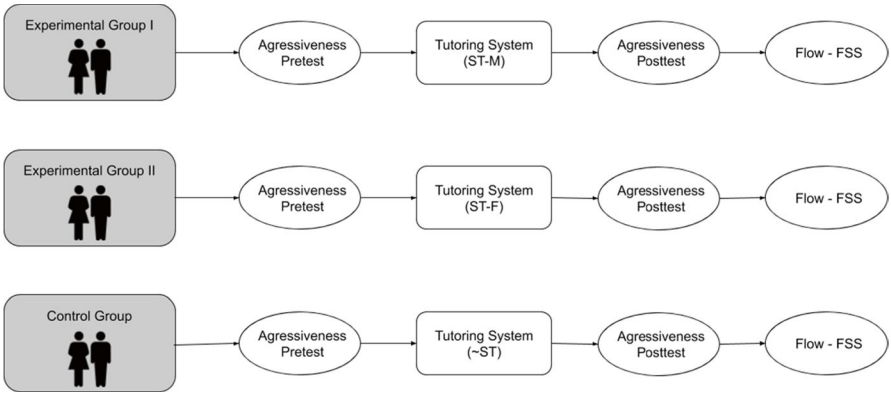

Fig. 2 Study Overview

$H_{1.c}$ Flow(control < aem = aef): The flow level is different in the control group when compared with the flow level of aem and aef. The flow level between aem and aef is equal;

H2: Participants allocated in gamified environments with stereotype threats will show different results on measures of aggressiveness, performance, and flow, as defined below:

$H_{2.a}$ Aggressiveness Level (threatening environment > control > non-threatening environment): Participants' aggressiveness levels in the threatening environment will be higher than the participants' aggressiveness levels in the control environment and the non-threatening environment. Participants in the non-threatening environment will have a lower level of aggressiveness than participants in the control environment;
$H_{2.b}$ Performance(non-threatening environment > control > threatening environment): The participants' performance in the non-threatening environment will be higher than the participants' performance in the control and threatening environments. Participants in the threatening environment will have lower performance than the participants in the control environment;
$H_{2.c}$ Flow(non-threatening environment > control > threatening environment): The participants' flow level in the non-threatening environment will be higher than the participants' flow level in the control and threatening environments. Participants' flow level in the threatening environment will have a lower flow level than participants in the control environment;

H3: Performance and flow measures will be mediated by the participants' level of aggressiveness;

Table 1 Subjects' Allocation: Gender X Platform

Contingency Table				
Gender	Control	AEF	AEM	Total
Female	23	19	21	63
Male	27	25	28	80
Total	50	44	49	143

2.6 Method

2.6.1 Subjects

The subjects of the study were randomly allocated to the experimental conditions, according to Table 1. The total of participants was 150, but we identified 143 valid responses because one participant refused to answer the gender question, and six did not answer the pre-aggression level.

The experiment was published on mailing lists in science, mathematics, engineering and psychology courses. Once they accessed the platform, participants read the consent form, and indicated their age and gender. Participants' ages averaged 18-24 years, followed by participants with a mean age between 25-34 years. The platform allocated participants randomly, with a balance of number of participants for homogeneity in their respective groups to avoid unequal analysis, for further analysis of these participants considering the next age groups: 18-24, 25-34, 35-44, 45-54. The groups of participants were later created to analyze the data from the evaluation instruments. For ethical reasons, the demographic information of the participants was anonymized to ensure total privacy in the collection of information and future statistical analysis.

The participants' distribution in the table shows a balance in the allocation to the six experimental conditions, considering the Bayesian factor ($BF_{10} = 0{,}069$) or chi-square $x^2(2) = 0{,}119$, $p = 0{,}942$).

3 Procedure

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

3.1 Gamified educational environments

The tool used was proposed by Albuquerque et al. (2017) and adapted for this experiment. In this tool is implemented the following gamification elements: points, badges, ranking, avatars, and trophies. These gamification elements were implemented in order to improve the users' experience and increase their attention and focus when performing the activities proposed on the platform.

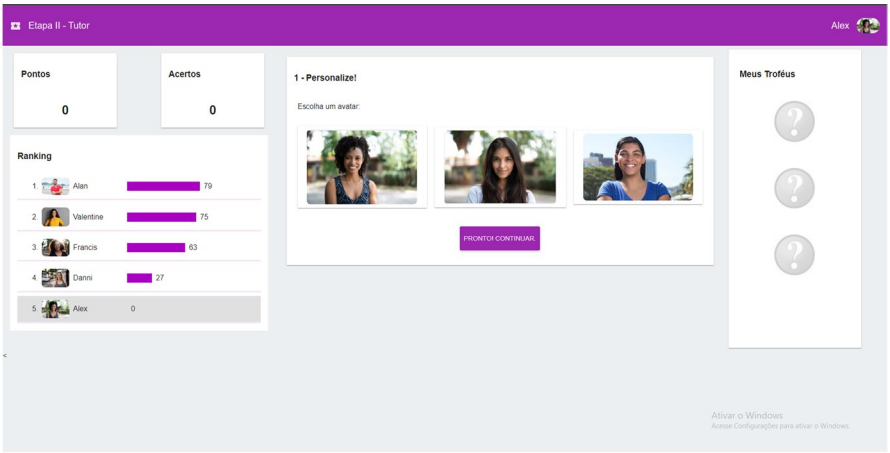

Fig. 3 Female Environment

Besides containing several graphical attributes that facilitate the users' engagement and learning, the platform provides its interface with different colors that were chosen according to the experimental groups: one female, one male, and one neutral. These different interfaces were randomly generated for each person who was going to use the tool, enabling different experiences within the same platform.

The choice of colors implemented in the different interfaces of the platform was based on the research (Anya C.H.Y., 2007). This study indicated which shades each gender felt most represented. As result, the three versions of the system: (i) stereotyped female environment (Fig. 3); (ii) stereotyped male environment (Fig. 4); and (iii) neutral environment (control) (Fig. 5). It is important to highlight that all the gamification elements of the platform were composed according to the colors of the interface in which they were located.

Given the literature discussed, the choices of alteration, in terms of colors and graphic elements inserted. The blue color choice for the stereotyped Male Environment gives hints of color with masculinity bias (Baliscei, 2020; Cohen, 2013), as well as for the pink color inserted in some elements of the stereotyped Female Environment (Baliscei, 2020). In aspects of biology, studies show that social contact, or even for genetic reasons, the effects that shape the tendency of adults to segregate their color preferences by gender. The gendered context of adulthood also matters, especially when women have children (Cohen, 2013; Del Giudice, 2012). Finally, other studies discuss and reinforce that the preference for these color preferences about gender happens through gender stereotypes and status differences between men and women (Jonauskaite et al., 2019).

Through this perception and the evidence found in the literature, this proposal uses the spectra of colors Blue, Gray, and Purple, as shown in the figures. The elements and gender dominance were also incorporated into the design of educational technologies.

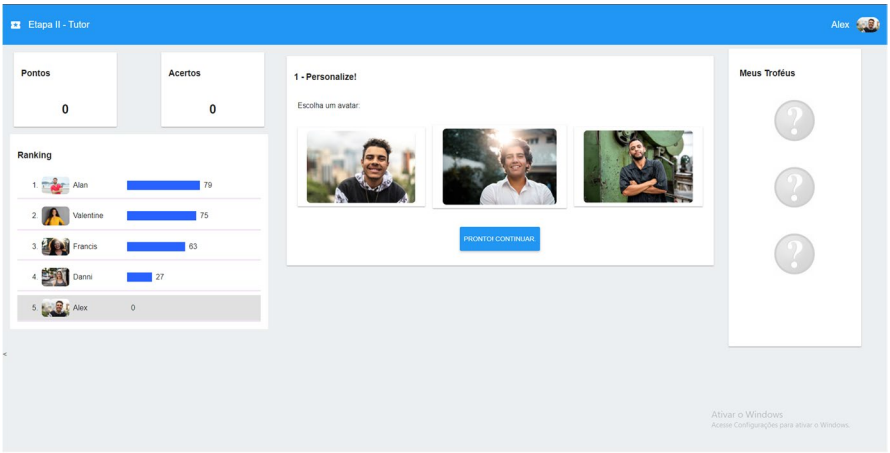

Fig. 4 Male Environment

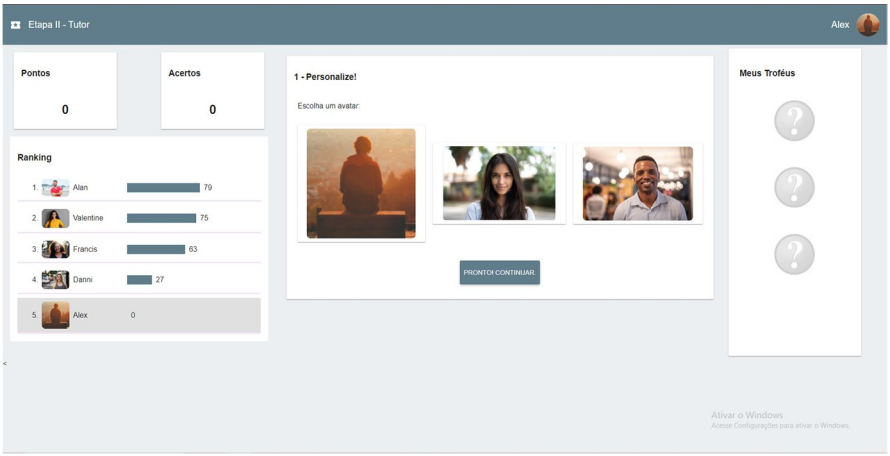

Fig. 5 Control Environment

Although they are subtle situations, the changes of each educational environment had the character of subtlety, since such elements, as well as colors, are essential for stereotyped environment representation with regard to technology design. In addition, educational environments were the closest to the educational reality. Traditionally, colors are used in various areas (marketing, advertising) as gender stereotypes (Yang and Li, 2016).

- **Female Environment**: The gamified tutor system with female gender stereotypes presents a predominantly purple color, with rankings that present only female gender participants, in addition to purple colored badges and avatars with only female illustrations;
- **Male Environment**: The gamified tutor system with the stereotype of the male gender features a predominant color is blue, with rankings that feature only male

gender participants, in addition to colored badges in blue and avatars with only male illustrations;

- **Control Environment**: The gamified tutor system with neutral attributes has a gray color, with rankings that feature participants of both gender, male and female, as well as avatars with male and female illustrations. This version was considered the control of this experiment.

In addition, the platform was adopted and used only for the experimental character. The participants of this experiment marked their consent to never having participated in this or other similar research. Thus, with guarantees unique subjects for the experiment. This version was developed using Python and AngularJS programming languages.

The following game elements are based on the components that are instances of the dynamics and mechanics observed in the gamified environments presented. Points are a quantitative metric that measures users' performance. As the participants answer questions, points could be removed or added depending on their response. The leaderboard is used to reveal how the student is doing compared to other participants on the platform. This element also shows how the student is advancing regarding the mastery of the contents. Avatar is a visual representation of the users' characters, and the users choose the character they are most familiar with or interested in. The avatars of the platform were redesigned, so that elements were used that made the individual who is using the platform feel more comfortable and represented. The avatars of the platform were redesigned in order to make the individual who is using the platform feel more comfortable and represented.

Points are a quantitative measure representing user performance on target behaviors to increase engagement. On the platform it was provided as follows: Every user starts with 0 points, in choosing the avatar he receives 10 points, and for each correct answer, 5 more points are attributed to his score, and for each error, 5 points are taken from his punctuation. Badges, or trophies, are awarded to the user when he achieves an achievement of 25 or 50 points. Avatar is the visual representation of a user's character, he chooses the one he is most familiar with, or has the most preference. The platform avatars were reformulated, so that elements were used that would make the individual who is using the platform feel more comfortable and represented. The colors that were used were related to Anya C.H.Y. (2007) which indicates which shades each genre has as a preference, and feels more represented.

3.2 Instruments

To measure the dependent variables, we used three scales: a scale to measure aggressiveness, a scale to measure performance, and a flow measure. The aggression questionnaire has 4 scales: Physical Aggression, Verbal Aggression, Anger, and Hostility. The final aggression score is a mean between the four scales. The flow scale is an instrument to measure a flow state of a user.

The degree of aggressiveness was measured in two moments using the same instrument. A pre-test and a post-test were answered by the participants, with 36 questions each based on Buss and Perry (1992) with four scales: Physical Aggression, Verbal Aggression, Anger, and Hostility. The application of a Bayesian t-test for independent measures showed that the mean of male and female participants did not differ significantly, both in the pre-test ($BF_{10} = 0.219$), and in the post-test ($BF_{10} = 0.232$) (t(141) = 0.650 , $p = 0.516$ e t(141) = 0.739, $p = 0.461$). Four outliers were identified in the pre-test and three in the post-test, but the scores were adjusted based on the nearest value. For the transformation of the outliers, according to Tabachnick et al. (2007) that if no substantial reason is given for the transformation of the outliers, they must be transformed. As there were only six cases and the modifications did not introduce substantial changes in the distribution of results and did not violate statistical principles, the most conservative solution of equating each outlier to the nearest non-outlier immediate value was adopted (Tabachnick et al., 2007). We measured the intellectual performance using an instrument, adapted from (Albuquerque et al., 2017), for the experiment (logic test) to measure the performance task, consisting of 20 items. The minimum score of the questionnaire was 0, and the maximum score reached was 110 points, with a mean of 61.7 and a standard deviation of 29.4. The distribution of results was normal, and we identified no outliers.

The flow measure was obtained using a questionnaire containing 26 questions according to Bittencourt et al. (2021). The validation results for this scale indicate good fit (X2-df = 2.94, CFI = 0.98, TLI = 0.97 and RMSEA = 0.053). The minimum value was 189.0, and the maximum value was 442.0, with a mean of 318.6 and a standard deviation of 56.7. Six outliers were identified, whose scores were also adjusted. The distribution of the scores was normal, with no violations in terms of skewness and kurtosis.

The conduction of this experiment involves the following steps:

1. Answer and accept the terms of the study;
2. Answer the Aggressiveness Test - Pre-Aggressiveness Measure;
3. 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;
4. Choice of avatars, according to the generated platform;
5. Answer the performance activity - Logical Test;
6. Answer the Aggressiveness Test - Post-Aggressiveness Measure;
7. Answer the Flow Test.

4 Performed statistical analyses

According to Berger et al. (1994), Bayesian inference is a statistical methodology based on the definition of probability as a degree of information. The main characteristic is the ability to combine new evidence with previous knowledge through the use

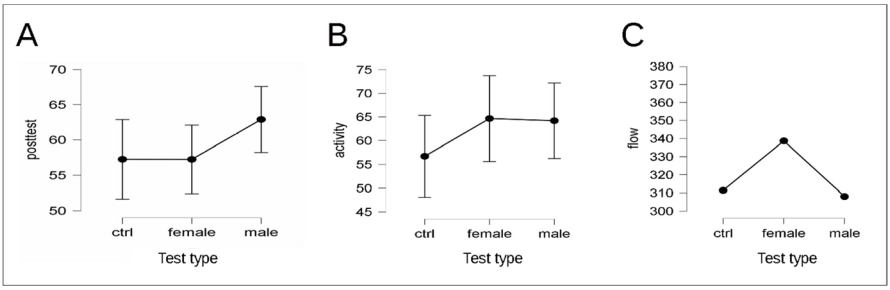

Fig. 6 Differences in the means of post-test aggressiveness (**A**), performance (**B**), and flow (**C**), as a function of the gamified environment

of the Bayes rule. In addition, four fundamental aspects that characterize the Bayesian approach to statistical inference can be identified: (i) A priori information (previous); (ii) Subjective Probability; (iii) Self consistency; and (iv) No "ad hoc" procedures.

Therefore, to test the hypothesis, we conducted a simple Bayesian ANOVA, Bayesian factorial ANOVA, and Bayesian factorial repeated-measures ANOVA. To measure correlation coefficients, was used Pearson's Correlation. Bayesian statistical conduction was used to determine how many times the null hypothesis is best based on the degree of probability. This is due to the interval nature of the data and the procedures of comparison between the experimental groups (control, male stereotyped environment, male stereotyped environment). To compare the effects between the groups, a Bayesian Post-hoc analysis was also conducted.

5 Results

In this section, all the statistical analyses performed in this study will be presented.

The first hypothesis tested aims to evaluate the isolated effects of gamified environments on the dependent variables. Initially, we tested the hypothesis that the type of environment caused differences in the level of aggressiveness, showing that an environment changed to meet gender-equity requirements would cause a different aggressiveness pattern when compared to a neutral environment. To test the hypothesis, we conducted a simple Bayesian ANOVA, with contrast different from 0, with values of -1.5 for the control and 1.0 for each of the stereotyped gamified environments. A contrast is a vector of weights with values that defines a specific comparison over means. They are used to testing more focused hypotheses than the overall omnibus test of the ANOVA (Hilton and Armstrong, 2006). The result was statistically significant ($t(141) = 6{,}589$, $p < .001$) and the means (observed in graph A plotted in Fig. 6), show that after conducting the experimental task, the male environment provided a detectable increase in aggressiveness. However, the same result did not occur with the stereotyped female environment, whose means did not differ from those recorded among participants allocated to the neutral environment.

In the test $H_{1.b}$ the same procedure of the previous analysis was adopted, adjusting the contrast (estimate of the true population value) values to -1,5 in the control

condition, and 1.0 for the two stereotyped environments. The result was statistically significant (t(141) = 2,118, p = .036), with the means (plotted in graph B of Fig. 6, suggesting that the stereotyped male and female environments increased the performance level of the participants compared to the results of the participants allocated to the control group. A similar procedure applied in the previous analysis was used to test H1c, and the results were also significant (t(141) = 2,063, p = .041). However, the means analysis (Fig. 6C), showed that only the stereotypical female environment provided a measurable increase in the flow level. In the stereotypical male environment, the score was close to those found among the participants allocated in the control group.

Although the previous results (Christy & Fox, 2014; Albuquerque et al., 2017; Jamieson & Harkins, 2012; Dasgupta et al., 2015; Cross et al., 2022) show that the gamified environment influences psychological variables as aggressiveness, intellectual performance, and flow level, they raise questions about the direction of these changes and the impact they may have on users' usability and performance in these systems. To address these questions, we analyzed the interaction between users' characteristics and these environments' particularities. Since we allocated male and female participants to stereotyped environments that were characterized as neutral, male and female, it was possible to conduct Bayesian factorial ANOVA to test the subsequent hypotheses.

To test $H_{2.a}$ we generated a Bayesian factorial repeated-measures ANOVA, with participant gender and environment type as IV and the two measures of aggressiveness, before and after, as DV. A model with a Bayesian factor equal to 84.5 was generated, a value that allows us to interpret that the hypothesis $H_{2.a}$ was 84 times better than the null hypothesis $H_{2.a}$ of absence of effect of the two variables on aggressiveness.

To present the results, we re-coded the participants' responses. We coded the responses of the males allocated to the stereotypical female environment as 1, and we coded the responses of the males allocated to the control and male environment conditions as 0. We adopted the same reasoning to re-code the female's answers. Therefore, we coded the responses of the females allocated to the male gamified environment as 1, and we coded the responses of other females as 0. This process allowed the inclusion of two new variables, threatened male and threatened female, whose scores are plotted in Fig. 7.

The scores show a clear tendency for participants with low and medium prior aggressiveness scores. These participants' post aggressiveness scores level increased when allocated to a threatening condition, an effect more noticeable among females. Among participants with high prior levels of aggressiveness, we identified that the threatening environment significantly reduced their post aggressiveness scores, mainly among males threatened by a female environment.

To test $H_{2.b}$, we conducted a Bayesian factorial ANOVA, with participant gender and the gamified environment as IVs and performance score as DV. The generated model favored the null hypothesis of no effects (BF01 = 151,321), which is a strong indicator that the interaction between participant gender and environment did not influence participants' performance.

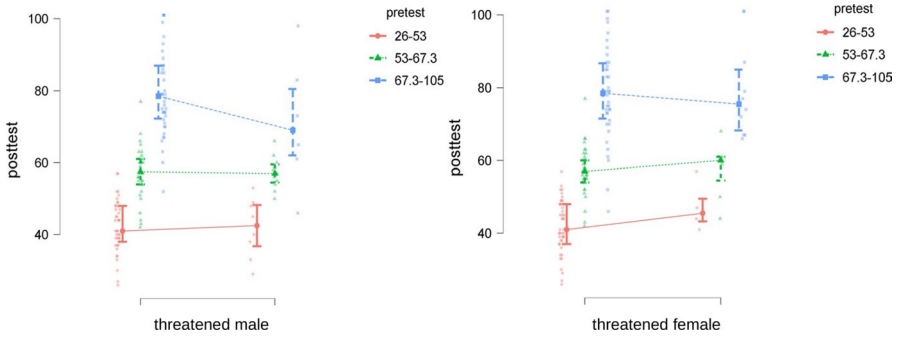

Fig. 7 Mean scores of aggressiveness in the post-test, by threat by threatening environments for men and women, segmented by aggression pre-test

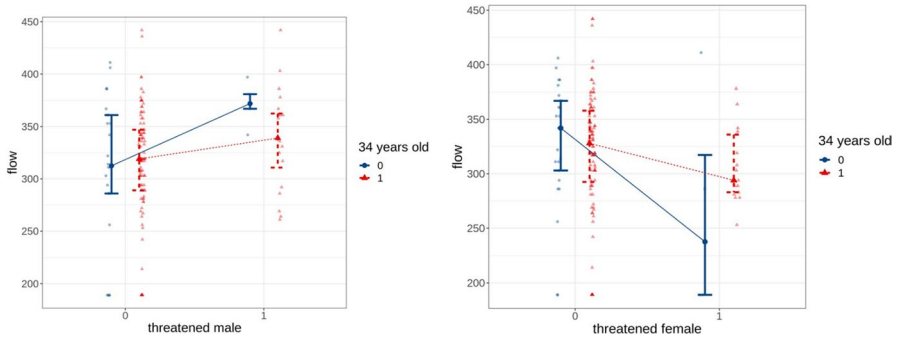

Fig. 8 Mean of flow scores by threatening environment for males and females and age

To test the hypothesis $H_{2.c}$, we conducted two Bayesian linear regressions, using the variables threatened male or threatened female and age above 34 years as predictors and the flow score as the DV. A regression model of good quality (BF10 = 4,272; R2 - 0,047), was generated for the threatened male, with the male threatening environment as the main predictor, while for the female threatening environment, the model was not significant (BF10 = 0,594; R2 - 0,018). The results, as plotted in Fig. 8, show that a threatening condition provided a significant increase in the flow level among males subjected to a threatening condition. This did not manifest itself with females, whom the patterns of graphical presentation of the results moved in the opposite direction.

To test the hypothesis H_3, we generated a correlation matrix among the variables flow, pre- and post-aggressiveness, and performance pattern, which allowed us to identify a correlation trend between flow and activity, as shown in Table 2

Based on this, variations in the aggressiveness's levels of the participants were identified, considering age groups (Pre and Pos Agre. $p < 0.001$ and r = 0.961; Flow and Performance $p = 0.091$ and r = 0.141). Therefore, the participants

Table 2 Correlation coefficients between the variables flow, performance, pre-test and post-test aggressiveness

Pearson's correlation					
Variable	Metric	Pre Agre.	Pos Agre.	Performance	Flow
Pre Agre.	Pearson's r	-	-	-	-
	p-value	-			
Pos Agre.	Pearson's r	0.961	-	-	-
	p-value	< 0.001	-	-	-
Performance	Pearson's r	0.042	0.052	-	-
	p-value	0.620	0.537	-	-
Flow	Pearson's r	-0.083	-0.104	0.141	-
	p-value	0.321	0.213	0.091	-

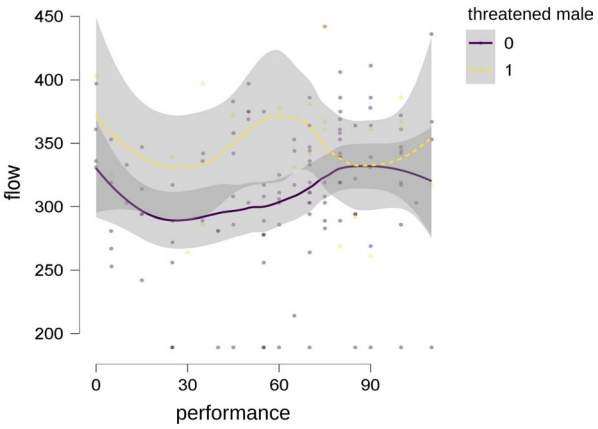

Fig. 9 Effects of the variables of performance and threatened male on the flow score

were divided into groups corresponding to their age. The group with the greatest variation in pre- and post-aggressiveness was the group with 34 years.

Posteriorly, we conducted a linear regression backward method with the variable flow as a criterion and the variables pre-aggressiveness, post-aggressiveness, performance, threatened man, threatened woman, and age up to 34 years as predictor variables. This allowed us to generate a model (adjusted r2 = 0,048; F (5,143) = 2,448, p = 0,037) with two predictors, threatened male and age up to 34 years, whose effects on the flow score can be seen in Fig. 9.

Figure 9 indicates representative oscillations in the relation between the lines, being identifiable as a marked distance when the participant obtains a medium score, in which the flow increases. Moreover, there are also notable marked approximations in the extreme positions of the graph, particularly when the score is from medium to high.

6 Discussion and conclusion

The results reported show that the male environment provided a detectable increase in the level of aggressiveness, and that the stereotypical male and female environments increased the level of performance of the participants. The literature may explain this phenomenon as the effect of novelty for the students, as a rewarding mechanism when performing the task proposed by the educational setting (Christy & Fox, 2014; Albuquerque et al., 2017). We also verified that the threatening condition provided a significant increase in the flow level among males submitted to a threatening condition, which did not manifest in the case of females.

The stereotyped gamified environments showed significant effects on the levels of aggression. This effect, generated by the stereotype threat, is similar to that found in the study of Inzlicht and Kang (2010) that studied the effects of the stereotype threats on social identity while doing math tests. In addition, as a result, in a threatening male environment, increased aggressiveness' levels could be detected, and this insight could generate emotional and cognitive stress. In addition, the control version environment was a system in which everyone had similar opportunities, and the literature discusses this effect (Kizilcec et al., 2017).

In terms of color perception and the use of stereotyped design elements, the results show an understanding of the phenomenon of colors. The elements used, cultural aspects, and characteristics inherited from parents or people through social interaction can also be generalized in colors, according to the blue color strongly present in stereotyped technology for men. Therefore, the version of the stereotyped technology for men influenced the behavior of women by the presence of blue color in the design or by the dominance of the design elements as avatars inserted for the male gender.

It is also important to highlight that, instead of impairing the performance levels of participants exposed to threatening environments, according to Inzlicht and Kang (2010) and Steele (1988), this study found a significant increase in the level of performance of participants exposed to threatening environments when compared to the control environment, which is also an interesting indicator of the assumptions of the multi-threat Framework which confirms that the subjects, even though members of the identity group, experience the threat in different ways (Shapiro and Neuberg, 2007).

The flow measurement was only significant in the female stereotyped environment, which could indicate that the subject's identification factor with the environment, in the case of women and women, can contribute to the valuation of positive psychological variables, such as performance and flow. Although this identification factor is present in the study by Rosenberg-Kima et al. (2010), which found significant changes in the performance of female students who identified with the avatars used on online platforms, and with other graphic elements, such as animations of their gender, our study identified that the participants who identified themselves as women showed a difference in the level of flow not significant when compared to the neutral environment, while men, in the same female stereotyped environment, presented. What may have occurred in this case is somewhat similar to that indicated by some studies of the literature review of Pennington et al. (2016) that showed

some subjects have a decrease in performance and other psychological variables. The increase in the measure of the flow of males in the female stereotyped environment may have been due to the need to meet chauvinist cultural precepts rooted in the social imaginary, in which an ideal of superiority between men and women is presented (Christy and Fox, 2014). The literature also discusses that concern based on comparisons may influence individuals' behaviors. In this scenario, the control environment, which was free of dominance comparison, generated an effect more positive for men (Albuquerque et al., 2017).

In conclusion, the study shows different performance levels in stereotyped platforms. In addition, the logical test could be easy for all subjects. The logical domain could be used as a hard task. A male group is a group with little influence on stereotypes. Furthermore, with the female environment, it was possible to have the best flow levels, indicating a characteristic discussed in the literature such as the feeling of competition between men, and affinities for women. In addition, the impact of observing levels of aggression in students using educational technologies is immense. Especially in the current scenario provoked by the pandemic generated by Covid-19, with thousands of students inserted into the context of online education. Observing the effect of stereotyped environments on aggressiveness provides a path to research, especially for the construction of technologies with the lowest incidence of stereotypes. Furthermore, it is necessary to construct studies that are capable of detecting stereotypes that are embedded in online educational technologies. This study points out initial indicators of characteristics that are inserted in the design of educational technology that can trigger a state of stereotype threat.

6.1 Threats to validity

6.1.1 Threats to internal validity

The pandemic caused by COVID-19 may have influenced the way users responded to the platform tests, considering that it is a period that has caused psychological suffering, raising of stress levels, melancholy, and possibly aggressiveness in the population because of confinement and constant threat to life. Furthermore, it is also possible that the participant may have had a stressful and/or distressing experience before and/or during performing the tests, which may also cause a difference in the aggressiveness levels. The colors used to indicate the wrong answers may have aroused signs of aggressiveness in the participants, since the exposure to the color red may affect the level of aggressiveness, as indicated by Vakili et al. (2019). Moreover, the sounds used to indicate the wrong answers may have aroused signs of aggressiveness in the participants, since the sound may have acted as an irritability or discomfort factor (Anderson and Bushman, 2002). Finally, the fact that participants always lose points when they miss a question may have been considered as an irritability factor. The control environment' design may have influenced the results of the activities and the Flow. Knowing that a trend can trigger risks with statistical analysis of correlational trends, we consider it necessary to show the correlation

between flow and performance, since these results were presented as a trend, with a risk of false-positive.

6.1.2 Threats to construct validity

The Aggressiveness index may have been impaired because there was no randomization in the order of the items included in the pre-test and the post-test. This could allow users to answer the post-test in the same way as they answered the pre-test.

6.2 Limitations

We did not collect age group and level of education information from the participants. This implies that we could not determine which age group, and/or to which level of education, the increase in aggressiveness in gamified educational environments is more related.

6.3 Future works

Our future works include conducting a study, similar to this article, with a focus on specific age groups and specific levels of education to investigate to what extent the effect of stereotype threat on aggressiveness is linked to the students' performance in these conditions. In this future work, we are going to reassess the items related to aggressiveness, randomizing their order in the pre-test and post-test. Moreover, we are going to add, in the logical reasoning test, an alternative that does not imply a loss of points for the participants. In addition, explore different colors and gamified elements. Some colors could be explored considering age preferences between gender. It could be a future work.

Acknowledgements This study was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ). The authors were also supported by the University of São Paulo (USP), the Federal University of Alagoas (UFAL), the ICMC-USP Software Engineering Laboratory (LABES), and the NEES - Núcleo de Excelência em Tecnologias Sociais.

Data availability The datasets analysed during the current study are available in the Dataverse repository: https://doi.org/10.7910/DVN/YGL56B.

Declarations

Ethics approval This study received approval from the ethics committee under the number 44824621.1.0000.5013 submitted to the national platform of the Brazilian National Council of Health (referred to as Plataforma Brasil). It is worth noting that data collection and storage have followed strict guidelines of data protection. In addition, all collected data have been anonymized.

References

Aguilar, D.A.G., Therón, R., & Peñalvo, F.J.G. (2013). Reveal the relationships among students participation and their outcomes on e-learning environments: case study. In *2013 ieee 13th international conference on advanced learning technologies* (pp. 443–447).

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.

Anderson, C.A., & Bushman, B.J. (2002). Human aggression. *Annual Review of Psychology*, *53*(1), 27–51.

Anderson, J.R, Boyle, CF., & Reiser, B.J (1985). Intelligent tutoring systems. *Science*, *228*(4698), 456–462.

Anya C.H.Y. (2007). Biological components of sex differences in color preference. *Current Biology*, *17*, R623–R625.

Bailey, D.R, & Lee, A.R. (n.d.) Learning from experience in the midst of covid-19: Benefits, challenges, and strategies in online teaching.

Baliscei, J.P. (2020). Abordagem histórica e artística do uso das cores azul e rosa como pedagogias de gênero. *Revista Teias*, *21*, 223–244.

Berger, J.O, Moreno, E., Pericchi, L.R., Bayarri, M.J., Bernardo, J.M., Cano, J.A., De la Horra, J., Martín, J., Ríos-Insúa, D., Betrò, B., & et al (1994). An overview of robust bayesian analysis. *Test*, *3*(1), 5–124.

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). *PloS one*, *16*(7), e0253044.

Bittencourt, I.I., Isotani, S., Costa, E., & Mizoguchi, R. (2008). Research directions on semantic web and education. *Interdisciplinary Studies in Computer Science*, *19*(1), 60–67.

Buss, A.H, & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, *63*(3), 452.

Cetintas, S., Si, L., Xin, Y.P.P., & Hord, C. (2009). Automatic detection of off-task behaviors in intelligent tutoring systems with machine learning techniques. *IEEE Transactions on Learning Technologies*, *3*(3), 228–236.

Cheung, A.C.K., & Slavin, R.E (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in k-12 classrooms: A meta-analysis. *Educational research review*, *9*, 88–113.

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. Retrieved from https://www.sciencedirect.com/science/article/pii/S0360131514001195.

Clarke, S.G, & Haworth, J.T (1994). 'flow'experience in the daily lives of sixth-form college students. *British Journal of Psychology*, *85*(4), 511–523.

Cohen, P.N (2013). Children's gender and parents' color preferences. *Archives of sexual behavior*, *42*(3), 393–397.

Colley, A., & Comber, C. (2003). Age and gender differences in computer use and attitudes among secondary school students: what has changed? *Educational Research*, *45*(2), 155–165.

Cross, L., Kaye, L.K, Savostijanovs, J., McLatchie, N., Johnston, M., Whiteman, L., Mooney, R., & Atherton, G. (2022). Gendered violence and sexualized representations in video games:(lack of) effect on gender-related attitudes. *New Media & Society* 14614448221075736.

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

Daouk, Z., Bahous, R., & Bacha, N.N. (2016). Perceptions on the effectiveness of active learning strategies. *Journal of Applied Research in Higher Education*.

Dasgupta, N., Scircle, M.M., & Hunsinger, M. (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proceedings of the National Academy of Sciences*, *112* (16), 4988–4993.

De Sousa Borges, S., Durelli, V.H., Reis, H.M., & Isotani, S. (2014). A systematic mapping on gamification applied to education. In *Proceedings of the 29th annual ACM symposium on applied computing* (pp. 216–222).

Del Giudice, M. (2012). The twentieth century reversal of pink-blue gender coding: A scientific urban legend?. *Archives of Sexual Behavior*, *41*(6), 1321–1323.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification. using game-design elements in non-gaming contexts. In *CHI'11 extended abstracts on human factors in computing systems* (pp. 2425–2428).

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Educational Technology & Society*, *18*, 75–88.

dos Santos, W.O., Bittencourt, I.I., Isotani, S., Dermeval, D., Marques, L.B., & Silveira, Ismar Frango (2018). Flow theory to promote learning in educational systems: Is it really relevant? *Revista Brasileira de Informática na Educação*, *26*(02), 29.

Francescucci, A., Kellershohn, J., & Pyle, M.A. (2020). Using online class preparedness tools to improve student performance: The benefit of "all-in" engagement. *Journal of Management Education*, 1052562920960205.

Friedman, B., Kahn, P.H., & Borning, A. (2008). Value sensitive design and information systems. *The handbook of information and computer ethics*, 69–101.

Gómez Contreras, J.L. (2020). Gamificación en contextos educativos: analisis de aplicación en un programa de contaduría pública a distancia. *Revista Universidad y Empresa*, *22*(38), 8–39.

Guerra, P.B.d.C. (2002). Psicologia social dos estereótipos. *Psico-USF*, *7*, 239–240.

Hilton, A., & Armstrong, R.A (2006). Statnote 6: post-hoc anova tests. *Microbiologist*, *2006*, 34–36.

Hwang, G.-J. (2003). A conceptual map model for developing intelligent tutoring systems. *Computers & Education*, *40*(3), 217–235.

Inzlicht, M., & Kang, S.K (2010). Stereotype threat spillover: how coping with threats to social identity affects aggression, eating, decision making, and attention. *Journal of personality and social psychology*, *99*(3), 467.

Inzlicht, M., Tullett, A.M, Legault, L., & Kang, S.K (2011). Lingering effects: Stereotype threat hurts more than you think. *Social Issues and Policy Review*, *5*(1), 227–256.

Jamieson, J.P, & Harkins, S.G (2012). Distinguishing between the effects of stereotype priming and stereotype threat on math performance. *Group Processes & Intergroup Relations*, *15*(3), 291–304.

Jonauskaite, D., Dael, N., Chèvre, L., Althaus, B., Tremea, A., Charalambides, L., & Mohr, C. (2019). Pink for girls, red for boys, and blue for both genders: Colour preferences in children and adults. *Sex Roles*, *80* (9), 630–642.

Kantharia, M. (2020). Online school education in india during coronavirus pandemic: Benefits and challenges. *Research Journal of Humanities and Social Sciences*, *11*(2), 99–103.

Kappen, D.L, Mirza-Babaei, P., & Nacke, L.E (2019). Older adults' physical activity and exergames: a systematic review. *International Journal of Human–Computer Interaction*, *35*(2), 140–167.

Kaye, L.K, & Pennington, C.R (2016). "girls can't play": The effects of stereotype threat on females' gaming performance. *Computers in Human Behavior*, *59*, 202–209.

Kaye, L.K., Pennington, C.R., & McCann, J.J. (2018). Do casual gaming environments evoke stereotype threat? examining the effects of explicit priming and avatar gender. *Computers in Human Behavior*, *78*, 142–150.

Kizilcec, R.F., & Lee, H. (2020). Algorithmic fairness in education. arXiv:2007.05443 [cs.CY].

Kizilcec, R.F., Davis, G.M., & Cohen, G.L. (2017). Towards equal opportunities in moocs: affirmation reduces gender & social-class achievement gaps in china. In *Proceedings of the fourth (2017) ACM conference on learning@ scale* (pp. 121–130).

Kizilcec, R.F., Saltarelli, A.J, Reich, J., & Cohen, G.L (2017). Closing global achievement gaps in moocs. *Science*, *355*(6322), 251–252.

Laureano-Cruces, A.L., Sánchez-Guerrero, L., Velasco-Santos, P., Mora-Torres, M., & Ramírez-Rodríguez, J. (2016). Design of pedagogical agents: the learning styles, the emotions and the color. In *E-Learn: world conference on e-learning in corporate, government, healthcare, and higher education* (pp. 421–431).

Liu, S., Liu, P., Wang, M., & Zhang, B. (2021). Effectiveness of stereotype threat interventions: A meta-analytic review. *Journal of Applied Psychology*, *106*(6), 921.

Lopes, V., Medina, R., Bernardi, G., & Nunes, F. (2019). Um modelo conceitual para adaptação contínua de elementos de gamificação em ambientes educacionais.

Myers, S.A, & Rocca, K.A (2000). The relationship between perceived instructor communicator style, argumentativeness, and verbal aggressiveness. *Communication Research Reports*, *17*(1), 1–12.

Orji, R., Tondello, G.F, & Nacke, L.E (2018). Personalizing persuasive strategies in gameful systems to gamification user types. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–14).

Ortner, T., & Sieverding, M. (2008). Where are the gender differences? male priming boosts spatial skills in women. *Sex Roles*, *59*, 274–281.

Pedro, L., & Isotani, S. (2016). Explorando o impacto da gamificação na redução do gaming the system em um ambiente virtual de aprendizagem. *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, *5*(1), 81. Retrieved from https://brie.org/pub/index.php/wcbie/article/view/6912.

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

Riley, J., & Ward, K. (2017). Active learning, cooperative active learning, and passive learning methods in an accounting information systems course. *Issues in Accounting Education*, *32*(2), 1–16.

Rosenberg-Kima, R., Plant, A., Doerr, C., & Baylor, A. (2010). The influence of computer-based model's race and gender on female students' attitudes and beliefs towards engineering. *Journal of Engineering Education, 99*.

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, *32*(1), 77–112.

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

Sheldon, L. (2012). The multiplayer classroom: Designing coursework as a game. *Cengage Learning*, *16*, 1–117.

Snyder, C.R., Lopez, S.J., Snyder, C.R., & Lopez, S.J. (2009). Mindfulness, flow e espiritualidade: em busca das melhores experiências. *Psicologia positiva: uma abordagem científica e prática das qualidades humanas. Porto Alegre: Artmed, 222–238*.

Steele, C.M (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology*, *21*(2), 261–302.

Tabachnick, B.G, Fidell, L.S, & Ullman, J.B. (2007). *Using multivariate statistics* Vol. 5. Boston: Pearson.

Vakili, H., Niakan, M.H., & Najafi, N. (2019). The effect of classroom red walls on the students' aggression. *International Journal of School Health*, *6* (1), 1–4.

Yang, F., & Li, C. (2016). The color of gender stereotyping: The congruity effect of topic, color, and gender on health messages' persuasiveness in cyberspace. *Computers in Human Behavior*, *64*, 299–307.

Yücel, Y., & Rizvanoglu, K. (2019). Battling gender stereotypes: A user study of a code-learning game, "code combat, " with middle school children. *Compute Human Behavior*, *99*, 352–365.

Authors and Affiliations

Jário Santos[1] · Ester Andrade[2] · Kamila Benevides[2] · Kelly Silva[2] · João Nascimento[3] · Ig Bittencourt[2,3] · Marcos Pereira[4] · Sheyla Fernandes[3] · Seiji Isotani[1]

✉ Jário Santos
jariojj@usp.br

Ester Andrade
elpa@ic.ufal.br

Kamila Benevides
kab@ic.ufal.br

Kelly Silva
kbas@ic.ufal.br

João Nascimento
joao.nascimento@ip.ufal.br

Ig Bittencourt
ig.ibert@ic.ufal.br

Marcos Pereira
emanoel@ufba.br

Sheyla Fernandes
sheyla.fernandes@ip.ufal.br

Seiji Isotani
sisotani@icmc.usp.br

[1] Institute of Mathematics and Computational Sciences, University of São Paulo (USP), Avenida Trabalhador São-Carlense, 400 Centro, 13566-590 São Carlos, SP, Brazil

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

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

[4] Institute of Psychology, Federal University of Bahia (UFBA), Campus Universitário, São Lãzaro, Salvador, BA 41940-220, Brazil