



The effects of gender stereotype-based interfaces on users' flow experience and performance

Wilk Oliveira^{1,2} · Juho Hamari² · William Ferreira¹ · Armando M. Toda¹ ·
Paula T. Palomino^{1,4} · Julita Vassileva³ · Seiji Isotani^{1,5}

Received: 8 February 2022 / Revised: 29 September 2022 / Accepted: 20 October 2022 /

Published online: 14 December 2022

© The Author(s) 2022

Abstract

Despite recent advances in the personalization of education, it is still unknown how different kinds of personalization affect students' experiences. To advance this literature, in this article, we present an experimental study with 307 participants investigating the effects of gender stereotype-based interfaces (in terms of colors and avatars stereotypes) on users' flow experience (*i.e.*, challenge–skill balance, merging of action and awareness, clear goals, feedback, concentration, control, loss of self-consciousness, and *autotelic* experience), and performance in a gamified educational system. The main results indicate that gender stereotype-based interfaces affect users' action–awareness merging, however, do not affect users' performance and overall flow experience. We contribute with the basis for new studies and challenge through future research attempts.

Keywords Educational technologies · Gender stereotypes · Personalized education · Flow experience · Data-driven study

✉ Wilk Oliveira
wilk.oliveira@tuni.fi

Paula T. Palomino
paula.palomino@nees.ufal.br

- ¹ Institute of Mathematics and Computer Science, University of São Paulo, São Carlos, São Paulo, Brazil
- ² Gamification Group, Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland
- ³ Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada
- ⁴ Computing Institute, Federal University of Alagoas, Maceió, Brazil
- ⁵ Graduate School of Education, Harvard University, Cambridge, USA

Introduction

In recent years, there have been a substantial growth of studies investigating the effects of personalized gamified educational systems in terms of students' perception (*e.g.*, students' motivation, performance, and flow experience) (Tondello et al., 2017; Toda et al., 2019b; Hallifax et al., 2019b; Oliveira et al., 2020). The results of these studies are contradictory (Klock et al., 2020). Some studies identified that the personalized gamified educational systems provided positive results [*e.g.*, improving motivation and collaboration (Vidergor, 2021), competence and satisfaction (Sailer et al., 2017), interaction time (Lavoué et al., 2018)]. Meanwhile, other studies show that personalized gamified educational systems provided neutral/negative results [*e.g.*, low motivation (Khoshkangini et al., 2017) or no significant difference in terms of flow experience (Oliveira et al., 2020)].

Especially in education, one of the most important and investigated phenomena is the flow experience (Oliveira et al., 2018), which is a feeling of deep engagement that a person can achieve during a given activity (Csikszentmihalyi, 2000). At the same time, flow experience is highly related to the learning experience (Csikszentmihalyi, 2014a). This means that when a student achieves the flow experience in an educational system, they will also be washed into a deep learning experience (Csikszentmihalyi, 2014a; Erhel & Jamet, 2019; Buil et al., 2019), the reason why it has been widely investigated in recent studies in the field of education (Hassan et al., 2020; Yoshida et al., 2013; Rodríguez-Ardura & Meseguer-Artola, 2019). This also indicates that when the personalization of the educational system is not done well and users need to interact in a system with stereotype threats, the flow experience can be low, negatively affecting the learning experience (Oliveira et al., 2021). However, even with this growing interest, the effects of personalization and gender stereotypes threats on students' flow experience are still poorly understood and the results differ between studies (both, positive and negative results) (Oliveira et al., 2018; Hallifax et al., 2019a; Oliveira et al., 2021).

One of the hypotheses for these contradictory results is that, in general, the studies are dedicated to analyzing only the effects of the gamification elements (isolated in or group) in the students' perception according to their gamer type/user type or demographic information (Koivisto & Hamari, 2019; Rapp et al., 2019; Bai et al., 2020), forgetting, therefore, other highly important aspects, such as gender stereotype threats [*e.g.*, a situation raises concern that one will be judged in terms of group stereotypes (Steele, 2011)] (Albuquerque et al., 2017; Oyibo & Vassileva, 2020). In particular, these aspects can be important for the personalization of gamified educational systems, since from older theoretical studies (Deaux & Lewis, 1984; Greene & Gynther, 1995) to more recent empirical studies (Dorji et al., 2015; Carrasco et al., 2017; Albuquerque et al., 2017; Oyibo & Vassileva, 2020; Santos et al., 2022), results demonstrate that stereotyped features such as colors and avatars (with specific gender stereotype characteristics) can positively or negatively affect users' experience, depending on their gender, for example, by increasing or decreasing concentration or anxiety (Albuquerque et al., 2017; Kollmayer et al., 2018; Komalawardhana & Panjaburee, 2018).

To meet this challenge, we present in this article the results of an experimental study ($N = 307$) analyzing the effects of gender stereotype-based interfaces on users' flow experience and performance in a gamified educational system. In this study, we seek to answer the following research questions: "Do gender stereotype-based interfaces affect the users' flow experience in gamified educational systems?" and "Do gender stereotype-based interfaces affect the users' performance in gamified educational systems?" By answering these research questions, we move toward the resolution of a global challenge, identifying whether gender stereotype-based interfaces in gamified educational systems can affect the users' flow experience and performance.

The main results indicate that the gender-stereotyped gamified educational system affected the users' action–awareness merging (*i.e.*, one of the flow experience dimensions), however, does not affect users' performance and the overall users' flow experience. Thus, our results allow us to advance the state-of-the-art, providing a basis for new studies and challenging the state-of-the-art for more thorough future research attempts. Our study, therefore, contributes to the areas of educational technologies, gamification, and gender studies, through insights related to the design of gender-based personalized gamified education systems.

Background

In this section, we explain the main contents labeled in this article (*i.e.*, Gender stereotype in gamified educational systems and Flow Theory). At the same time, in this section, we present and compare the main related publications.

Gender stereotype in gamified educational systems

Gamification *i.e.*, the idea of "transforming systems, services, and activities to better afford similar motivational benefits as games often do" (Koivisto & Hamari, 2019; Hamari, 2019), when applied in education, often aims to improve the students' experience (*e.g.*, students' engagement, motivation, and flow) in educational environments (Oliveira & Bittencourt, 2019; Toda et al., 2019a; Janelli & Lipnevich, 2021). The idea of personalized gamified educational systems arose from the fact that people have different personalities and characteristics (*e.g.*, gender, age, gamer/user type) that may influence their preferences regarding gamification design (Tuunanen & Hamari, 2012; Belk et al., 2014; Vail et al., 2015). Thus, to provide a more suitable gamification design it is important to personalize it to the user's preferences (Oliveira & Bittencourt, 2019; Masthoff & Vassileva, 2015). Then, personalized gamified educational systems is a novel research field (Hallifax et al., 2019b), with few empirical studies conducted (Klock et al., 2020) and with conflicting results in terms of the student experience (Hallifax et al., 2019a), with room for further research that can further investigate different effects of personalized gamified educational systems on students' experience.

In recent years, within the personalization of gamification domain, few studies have also aligned discussions related to gender-based personalization (*i.e.*,

gender-based personalization strategies, according to the gender¹ self-reported by users) and gender stereotype threats (*i.e.*, situation in which people of a certain gender do not feel represented in a certain system (Steele, 2011; Albuquerque et al., 2017; Santos et al., 2022)) as a determining factor to affect the students' experience (Albuquerque et al., 2017; Göbl et al., 2021; Denden et al., 2021). Especially in gamified education, this is because, commonly, gamified educational systems are designed/implemented as a male-stereotyped system (*e.g.*, with only male-stereotyped avatars, male-stereotyped colors, or male-stereotyped language) (Wanner et al., 2020).

This tendency tends to negatively affect the perception of certain users (Oyibo et al., 2017), who may, for example, have a high level of anxiety when using a stereotyped interface for another gender. At the same time, this can lead to students giving up on using a system (Albuquerque et al., 2017). Despite these theoretical assumptions, there are still a lack of empirical/experimental studies that investigate the effects of gender stereotype-based interfaces on the students' experience (*e.g.*, flow experience) and performance.

Flow theory

The Flow Theory was proposed in the 1970s by Csikszentmihalyi and Csikszentmihalyi (1975) and represents an experience of deep engagement in certain activities (Csikszentmihalyi, 1997a). To achieve a flow experience, one has to go through nine different dimensions which when achieved together represent the so-called flow experience: (i) challenge–skill balance; (ii) action–awareness merging; (iii) clear goals; (iv) unambiguous feedback; (v) total concentration on the task at hand; (vi) sense of control; (vii) loss of self-consciousness; (viii) transformation of time; and (ix) *autotelic* experience (Csikszentmihalyi, 2000).

The idea of challenge–skill balance can be defined in a personal way, separated from any structures of activity, and the perception of the defined challenge is critical to flow occurring Jackson and Marsh (1996), Jackson and Eklund (2002), and according to Csikszentmihalyi and Csikszentmihalyi (1975), when in flow, a dynamic balance exists between challenges and skills. At the same time, the idea of action–awareness merging, according to Jackson et al. (2011), is a sense of effortlessness and spontaneity, associated with the flow dimension of action–awareness merging (further details next), coming about through a total absorption in what one is doing.

The clear goals dimension is a necessary part of achieving something worthwhile in any endeavor and the focus that goals provide to actions also means that they are an integral component of the flow experience (Jackson et al., 2011; Jackson & Eklund, 2002). In this way, the unambiguous feedback dimension is closely associated with clear goals and is represented as the processing of how performance is progressing in relation to these goals (Jackson et al., 2011).

¹ Considering gender different from biological sex (De Lauretis, 1987).

The total concentration on the task at hand is a dimension that represents a person totally connected to the task in which one is engaged and optimizes the flow state and this *connectedness* relies on a present-centered focus (flow resides in being in the present moment, rather than in the past or future) (Jackson et al., 2011; Jackson & Eklund, 2002). At the same time, having the experience of total control (*i.e.*, sense of control dimension) is likely to move an individual away from the experience of flow and into relaxation or boredom.

The loss of self-consciousness dimension is related to liberating to be free of the voice within our head that questions whether we are living up to self- or other imposed standards Jackson et al. (2011), while the transformation of time dimension can be represented as the intensity of focus may also contribute to perceptions of time slowing, with a feeling of having all the time in the world to execute a move, that is, in reality, time limited (Csikszentmihalyi & Csikszentmihalyi, 1975; Csikszentmihalyi, 2014b; Jackson et al., 2011; Jackson & Eklund, 2002).

Finally, the *autotelic* experience is often recognized as the most important flow experience dimension. Csikszentmihalyi (1997b) coined the term “*autotelic*” experience to describe the intrinsically rewarding experience that flow brings to the individual. It is generally after completing an activity, upon reflection, that the *autotelic* aspect of flow is realized and provides high motivation toward further involvement (Jackson et al., 2011).

Different studies conducted over several years show that this experience is highly linked to students’ learning experience because when in a flow experience in an educational activity, the learning experience is also good (Csikszentmihalyi, 2014b; Buil et al., 2019; Rodríguez-Ardura & Meseguer-Artola, 2019). Furthermore, recent studies have drawn attention to creating educational activities that may lead to the flow experience (Gao et al., 2019; Oliveira et al., 2019, 2020). At the same time, a system with stereotype threats can negatively affect the user experience (Dorji et al., 2015; Albuquerque et al., 2017; Santos et al., 2022), and if it negatively affects the flow experience, it can also harm the learning experience.

Given the direct relationship between the flow experience and the learning experience (Csikszentmihalyi, 2014a; Heutte et al., 2016), it is important to identify which gamification aspects (*e.g.*, colors, design dos elements, personalization) can lead students to a flow experience (Oliveira et al., 2020). Thus, one important topic within Flow Theory in educational technologies studies is to identify the gamified educational systems design aspects (*e.g.*, gender stereotype) that can lead to this experience (Kiili et al., 2012; Hsieh et al., 2016; Erhel & Jamet, 2019).

Related works

To identify the related works, we analyzed recent systematic literature reviews in the field of gamification (Koivisto & Hamari, 2019), personalized gamification (Halifax et al., 2019a; Rodrigues et al., 2020), and Flow Theory (Oliveira et al., 2018, 2021). We focus on presenting studies dealing with gender stereotype threats or Flow Theory in gamified educational systems. We also present studies regarding the effects of color temperature in users’ experience.

Concerning studies analyzing the effects of gender stereotype in gamified educational systems on participants' experience, Albuquerque et al. (2017) conducted an experimental research investigating if gender stereotype threat in a gamified educational system affects the users' anxiety and performance. They executed a three-stage survey where participants were asked about their anxiety after they used an online gamified educational system (to solve a logic quiz) and finally they were asked about their anxiety during the system usage (thus, comparing the students' anxiety before and during the system usage). One of their results indicated that the male-stereotyped system increased the females' anxiety (Albuquerque et al., 2017). This study opens space for conducting new studies investigating different aspects of user experiences in this kind of system (*e.g.*, concentration and flow).

Using the same system, Santos et al. (2022) investigated the effects of the stereotype threats using a quantitative experiment in three gamified environments (stereotypical male version, stereotypical female version, and control environment). The study was conducted with a sample size composed of 150 high school and undergraduate students. They identified that the participants randomly assigned to the male learning environment present an increase in aggressiveness level. Among other results, they also identified that the stereotypical male and female learning environments increased the participants' performance level (Santos et al., 2022).

Wanner et al. (2020) investigated how male priming in STEM subjects affects female emotions during gamification tasks. Female students who were primed with Anti-STEM achieved better results in the programming gamified task. It could be observed that the girls enjoyed the gamification tasks (Wanner et al., 2020). The result confirms that gamified activities can affect participants differently based on gender.

Denden et al. (2021) examined the effect of gender and personality differences on students' perception of gamification in education. In a study with 189 undergraduate students, they observed that gender and personality can affect students' perception of specific game elements. Especially, females are more likely to find feedback useful than males. They also observed that gender moderates the effect of personality on students' perception of the implemented game elements (Denden et al., 2021). The results of Denden et al. (2021) also demonstrate the importance of analyzing different design aspects that affect the students' experience.

Other studies have been devoted to analyzing the effects of personalizing educational games and gamified educational systems on participants' flow experience. Erhel and Jamet (2019) conducted a study investigating the relationship between the flow state and learning within an educational game. Their results indicate a positive influence of participants' flow experience in their memorization and comprehension. The study, however, was carried out in the context of games (and not gamification) and did not consider any type of gender stereotype (Erhel & Jamet, 2019).

Oliveira et al. (2020) investigated the effects of a player type-based personalized gamified educational system in terms of students' flow experience. They executed an experimental study with a sample composed by 121 participants comparing a tailored with a counter-tailored version. It was not possible to notice a significant difference in the students' experience when using the system (Oliveira et al., 2020). These results contradicted some other similar studies and also draw attention to the

Table 1 Related works comparison

Studies	FE	ES	NSD	GPS	SS	MR
Albuquerque et al. (2017)	No	Yes	Yes	Yes	127	+A
Santos et al. (2022)	Yes	Yes	Yes	Yes	150	+AA
Wanner et al. (2020)	No	Yes	Yes	Yes	NA	+S
Denden et al. (2021)	No	Yes	Yes	Yes	189	+P
Oliveira et al. (2020)	Yes	Yes	Yes	No	121	−+
Erhel and Jamet (2019)	Yes	Yes	Yes	No	109	+MC
Oyibo and Vassileva (2020)	No	Yes	Yes	No	323	+C
Our study	Yes	Yes	Yes	Yes	307	

FE analyzes the flow experience, *ES* is an experimental study, *NSD* uses non-simulated data, *GPS* uses a gender-stereotyped system, *SS* sample size, *MR* main result, *NA* not available, *+A* male-stereotype design increase participants anxiety, *+AA* male-stereotype design increase participants aggressiveness, *+P* gender and personality can affect users' perception, *+S* anti-STEMs positively affect female users, *−+* without significant results, *+MC* flow positively affects memorization and comprehension; *+C* colors affecting users' experience

importance of conducting new studies to investigate the flow experience in other types of personalization.

Studies have also investigated the effects of colors on users' experience. Oyibo and Vassileva (2020) conducted an empirical study to investigate the effects of color temperature and layout design in terms of technology acceptance and user experience (UX). They conducted an experiment with a sample composed of 323 respondents. Their finds suggest that the colors blue and green are more useful than orange and red in tourism websites (Oyibo & Vassileva, 2020). The study conducted by Oyibo and Vassileva (2020) focused on mobile interfaces for tourism applications, opening space for similar investigations in the field of educational systems.

In summary, the related studies are subdivided into (i) analyzing the effects of gender stereotypes on users' experience, (ii) analyzing the effects of personalized gamification (*e.g.*, user type-based personalization) on the users' flow experience, or (iii) analyzing the effects of using colors on the participants' experience. However, there are a lack of studies that investigate the effects of gender stereotype on users' flow experience and performance in a gamified educational system. Thus, as far as we know, our study is one of the first to investigate whether gender stereotype affects users' flow experience and performance in a gamified educational system. Table 1 presents a comparison between the related works.

Study design

This study aims to analyze the effects of gender stereotype-based interfaces on users' flow experience and performance in a gamified educational system. To conduct our experiment, the Goal/Question/Metric (GQM) framework was used. The GQM was

chosen because its efficiency is sustained by (i) define project goals; (ii) verify the goals and define how the study goals will be achieved; and (iii) provide an adequate interpretation of the results based on the established goals (Caldiera & Rombach, 1994). At the same time, this framework is used and recommended for experimental studies in different areas (Van Solingen et al., 2002; Aljedaibi & Khamis, 2019; Tuah & Wills, 2020).

Hypotheses definition

Over the years, several studies have shown that human beings have different perceptions/preferences regarding colors (Hallock, 2003; Karniol, 2011; Beigpour & Pedersen, 2015), as well as different perceptions/preferences regarding gender stereotypes (Basow, 1992; Deaux & Lewis, 1984). Studies conducted by Greene and Gynther (1995), Coursaris et al. (2008) and others, for example, identified that males tend to prefer and feel better in environments with blue colors, while girls feel better in environments of pink color, at the same time that gray color is a neutral color for both males and females. Other recent studies also show that the stereotypes of personalization of systems can also influence (positively or negatively) in the users' perception (Albuquerque et al., 2017; Oyibo & Vassileva, 2020).

Studies also show that possibly, people tend to have different preferences for avatars, according to the stereotype of those avatars (*e.g.*, avatars with characteristics related to their gender) (Albuquerque et al., 2017; Carrasco et al., 2017). Different studies have shown that when someone needs to use a stereotyped system for another gender, they tend to feel uncomfortable in the system, for example, having a significant increase in anxiety (Albuquerque et al., 2017). At the same time, other studies highlight that these characteristics can directly influence students' learning/performance when interacting with educational systems (Orji et al., 2013, 2014). Thus, we hypothesized that *gender stereotype-based interfaces affect the students' flow experience in a gamified educational system* (H1), as well as *gender stereotype-based interfaces affect the student's performance in a gamified educational system* (H2).

Participants

Our participants were 307 (173 self-reported as male and 134 self-reported as female) subjects with an average age of 24 years old ($SD = 3262$), from 26 different countries. Our participants were recruited through the Amazon Mechanical Turk (MTurk)² and Prolific platform³, which are crowdsourcing marketplace services highly used to conduct human-based experiments (Paolacci et al., 2010; Palan & Schitter, 2018; Orji et al., 2014; Hallifax et al., 2019b). Each participant from MTurk received 25 cents, while each participant from Prolific received 0.63 £ for

² <https://www.mturk.com/>.

³ <https://prolific.co/>.

Table 2 Participants per country

Country	Number of participants
Others	1 each
China	2
Chile and Greece	4
The UK and Brazil	5
Spain	9
Italy	11
Mexico	17
Poland	20
Portugal	29
India	53
The USA	137

Others—Canada, Romania, Estonia, France, Germany, Hungary, India, Ireland, Latvia, Morocco, New Zealand, Panama, Slovenia, and Turkey and Trinidad

their participation. For MTurk participants, the value was decided based on a pilot study, conducted with 10 participants, where they were asked about the value they considered fair for participation in the experiment. For Prolific participants, the value is automatically defined by the platform, based on the average response time of the participants. We also invited totally voluntary participants from email lists and social networks (only 5 were volunteers). This data collection strategy aimed to ensure maximum heterogeneity in our sample. Table 2 describes our sample.

The study is suitable for different criteria for sample size validation. According to Bentler and Chou (1987) it is necessary to have at least five participants for each construct measured (in our study, we have nine constructs and 307 participants, thus ≈ 34 for each construct). Hair et al. (1998) suggests the same rule for factor analyses. Loehlin (1998) suggests that at least 100 participants are required for a complete sample size.

Materials

To identify the participants' experience, we used the open-source-gamified educational system called "Removed for anonymous review," which was implemented to help participants in studying different educational contents in a simple and fun way, through a simple design and the use of gamification elements (Albuquerque et al., 2017). Initially, the system allows the users to choose an avatar to represent themselves and personalize their experience in the system (see Fig. 5a in Appendix 1). Then the users can begin to study a particular subject (*e.g.*, equation) and answer questions about that subject (see Fig. 5b in Appendix 1). The system was chosen because it uses the most common gamification elements in the field of education (*i.e.*, points, badges, ranking, levels, progress bar, and avatar) according to different

recent literature reviews (Hamari et al., 2014; Sousa Borges et al., 2014; Nah et al., 2014; Koivisto & Hamari, 2019; Oliveira et al., 2021).

The system has three versions. Figure 5a and b (in Appendix 1) shows the neutral version of the system. In the neutral version, only neutral colors (which are usually associated with no gender (Greene & Gynther, 1995; Karniol, 2011)) are used. In this version, there are avatars with male, female, and neutral stereotypes to be chosen by users. Similarly, gamification elements are presented in neutral colors. Figure 5c and d (in Appendix 1) shows the male-stereotype version of the system. In this version, the color blue [the main color associated with the male gender (Greene & Gynther, 1995; Karniol, 2011)] was used. In choosing avatars, only characters with male stereotypes were used. Finally, Fig. 5e and f (see Appendix 1) show the female-stereotype version of the system. For this version, the color lilac [the main color associated with the female gender (Greene & Gynther, 1995; Karniol, 2011)] was used. In choosing avatars, only characters with female stereotypes were used. The system personalization was based on different studies on gender color design stereotypes (Greene & Gynther, 1995; Hallock, 2003; Hurlbert & Ling, 2007; Cour-saris et al., 2008; Karniol, 2011). These studies show that people prefer different colors according to their gender and indicate which colors are most associated with each gender.

To identify the flow experience of participants during the system usage was used the short flow state scale (short FSS) was developed and validated by Jackson and Eklund (2002). The scale was developed based on the nine original Csikszentmihalyi and Csikszentmihalyi (1975) Flow Theory dimensions. This scale was also validated for the gamification domain (*i.e.*, to be used in gamified systems) by Hamari and Koivisto (2014). For this study, we used the short FSS following the original “Manual for the Flow Scales” (Jackson et al., 2011), presenting the scale in a 5-point Likert Scale (Likert, 1932). This scale was chosen because according to the systematic literature review conducted by Oliveira et al. (2018), the scale uses the nine original flow experience dimensions proposed by Csikszentmihalyi (1997a) and is the most used scale in the field of educational technologies. We also included an “attention-check question” (*i.e.*, if you are filling out the form carefully, answer 4) to remove participants who were not paying attention when answering the scale. Appendix 2 presents the FSS, including information regarding how each flow experience dimension is measured in the scale. SPSS 27 software program was used to conduct the analysis.

Procedure

The study procedure was organized in three different steps (pilot studies demographic survey, system usage, and flow experience report):

- *Pilot studies* In this step, we conducted an initial study, where we evaluated the system used in this experiment with six Master of Business Administration (MBA) students (with voluntary participation in the study). In this pilot study, we applied the think-aloud protocol (Alhadreti & Mayhew, 2016) just to make an

initial qualitative analysis of the student's perception regarding the system personalization and to evaluate the possibilities of changes in the conduct of the main experiment. The think-aloud protocol was chosen to be used in this step because it is a widely used approach to user interface analysis, capable of perceiving details even with small samples (Charters, 2003; Alhadreti & Mayhew, 2016). Then, we conducted a second pilot with 10 participants in the MTurk. The sole purpose of this pilot study was to assess the fairest amount to pay for those participating in the main study.

- *Demographic survey* In this step, participants answered a demographic survey asking about their gender, age, academic/educational level, and birthplace (country). Especially, self-reported gender information was important to analyze the flow experience level according to the user gender and the system version used. These data were chosen because it is widely used in demographic studies in the field of educational technologies [e.g., (Orji, 2014; Hallifax et al., 2019b; Oliveira et al., 2020)]
- *System usage and flow experience report* In this step, participants used logic quizzes (within the gamified educational system) for about 25 min. Upon entering the system, a randomization algorithm automatically drew one of the system versions (male stereotyped, female stereotyped, or neutral) for each participant. Then, immediately after finishing the system usage, participants answered the FSS according to their experience during the system usage.

We initially collected and grouped the data from all responses. We removed responses of participants who missed the attention-checking question and participants who choose different options from males or females on the demographic survey. We received originally 330 responses, 17 were removed because answered wrong the “validation question.” No participant selected any different option of male or female and prefer not to inform the gender, which has been removed from the analysis. In our analysis, we considered three personalization settings (male stereotyped, female stereotyped, and neutral), which is our independent variable, as well as, the participant's performance and all flow experience dimensions (our dependent variables). In our study, we compared all possible personalization settings (*i.e.*, male using the neutral version (MD), female using the neutral version (FD), male using the male-stereotyped version (MM), male using the female-stereotyped version (MF), female using the female-stereotyped version (FF), and female using the male-stereotyped version (FM)).

Data analysis

To analyze the data, first, we compared the different flow experience dimensions [*i.e.*, challenge–skill balance; action–awareness merging; clear goals; unambiguous feedback; total concentration on the task at hand; sense of control; loss of self-consciousness; transformation of time; and *Autotelic* experience (Csikszentmihalyi, 2000)] in each group of participants and the overall flow experience [obtained by the FSS (Jackson & Eklund, 2002; Hamari & Koivisto, 2014)]. To calculate the

Table 3 Participants' flow experience analysis

	Flow	CSB	MMA	G	F	C	CTRL	LSC	T	A
Female users using the default system										
Mean	3.717	3.935	2.848	4.043	4.087	4.239	4.174	3.696	3.043	3.543
SD	0.577	0.942	1.160	0.806	0.855	0.937	0.789	1.365	1.250	1.155
Male users using the default system										
Mean	3.980	4.157	3.333	4.020	4.235	4.392	4.216	3.765	3.627	3.902
SD	0.610	0.801	1.166	0.939	0.854	0.743	0.870	1.214	1.283	0.869
Female users using the female-personalized system										
Mean	3.854	3.917	2.938	4.021	4.042	4.542	4.021	4.083	3.250	3.688
SD	0.707	0.909	1.313	0.989	0.889	0.735	1.051	1.017	1.346	0.939
Male users using the female-personalized system										
Mean	3.934	4.246	3.246	4.098	4.246	4.426	4.328	3.902	3.197	3.623
SD	0.624	0.739	1.363	0.694	0.782	0.799	0.824	1.183	1.252	1.104
Female users using the male-personalized system										
Mean	3.875	4.050	2.500	4.175	4.075	4.650	4.350	4.025	3.275	3.725
SD	0.640	0.999	1.323	0.803	0.905	0.527	1.014	1.351	1.581	1.162
Male users using the male-personalized system										
Mean	3.770	4.066	2.672	4.082	4.066	4.426	4.016	3.754	3.131	3.443
SD	0.584	0.921	1.155	0.874	0.847	0.778	1.063	1.196	1.138	1.017

CSB challenge–skill balance, *MMA* action–awareness merging, *G* clear goals, *F* unambiguous feedback, *C* total concentration on the task at hand, *CTRL* sense of control, *LSC* loss of self-consciousness, *T* transformation of time, *A* *autotelic* experience

participants' flow experience in each version, comparing its difference, we calculated the median, standard deviation (SD), and normality of the data for all flow experience dimensions.

As the population of each group was less than 50, following the recommendations of Wohlin et al. (2012), we decided to use the Shapiro–Wilk test to analyze the data normality (Shapiro & Wilk, 1965). After this, following the Shapiro–Wilk test results (where we detect that the data do not follow a normal distribution), we analyzed the variance between each flow experience dimension for each group previously defined. Then, to verify our hypotheses, considering that we cannot affirm that the data are within a normal distribution, following the recommendations of Wohlin et al. (2012), we performed the Kruskal–Wallis test.

Results

In Table 3, we present the descriptive statistics of the study, showing the mean and SD for all participants' flow experience dimensions in the different versions of the gamified educational system (including overall flow experience). Figure 1 presents a graphical comparison of the users' experience and Fig. 2 presents a graphical comparison of the users' performance during the system usage.

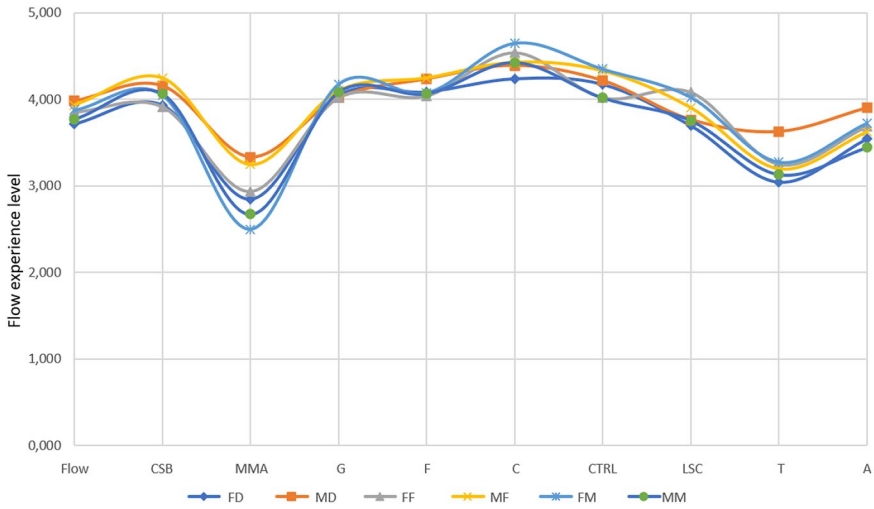


Fig. 1 Users’ flow experience analysis

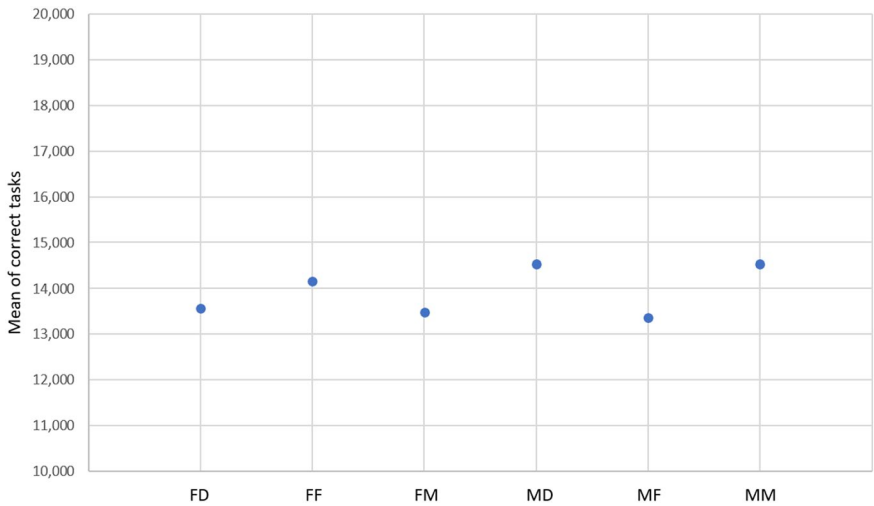


Fig. 2 Users’ performance analysis

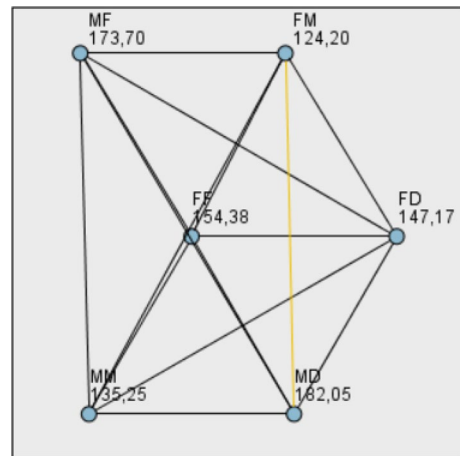
Table 4 presents the Kruskal–Wallis test, while Fig. 3 presents the pairwise comparisons for the Kruskal–Wallis. Results of Kruskal–Wallis test showed that gender stereotypes in gamified educational systems does not affect students’ performance and overall students’ flow experience, however, affect students’ action–awareness merging ($k = 16,393; p = 0.006$) that is one of the flow experience dimension. So, we carried out the post hoc tests for action–awareness merging. Table 5 presents the post hoc results and Figure 4 presents the independent samples Kruskal–Wallis test for the action–awareness merging.

Table 4 Kruskal–Wallis test

	CSB	MMA	G	F	C	CTRL	LSC	T	A	Flow	P
<i>k</i> value	5.21	16,393*	0.785	3.66	6.225	6.303	4.28	6.801	6.712	6.979	9.697
<i>p</i> value	0.390	0.006	0.978	0.599	0.285	0.287	0.51	0.236	0.243	0.222	0.084

k value Kruskal–Wallis test, *CSB* challenge–skill balance, *MMA* action–awareness merging, *G* clear goals, *F* unambiguous feedback, *C* total concentration on the task at hand, *CTRL* sense of control, *LSC* loss of self-consciousness, *T* transformation of time, *A* *autotelic* experience, *P* performance

* $p < 0.05$

Fig. 3 Pairwise comparison

Discussion

In this article, we present the results of an experiment analyzing the effects of gender stereotype-based interfaces on users' flow experience and performance in a gamified educational system. The main results indicate that gender-stereotyped interfaces do not affect users' performance and overall users' flow experience, however, affect users' action–awareness merging.

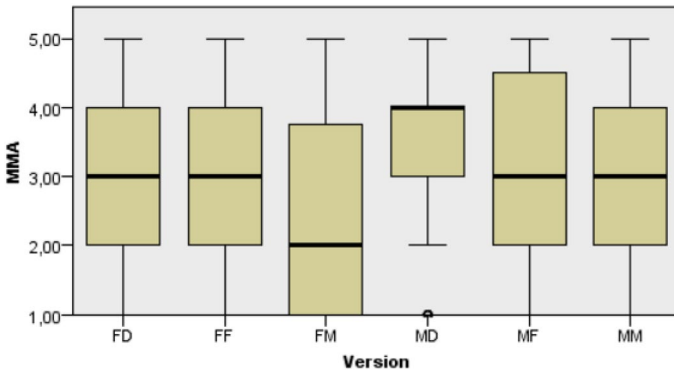
When analyzing the trend presented in Fig. 1 and Table 3, it is possible to notice a similarity in terms of the effects of stereotypes in the different flow experience dimensions. That is, when an experience was high or low in one of the system's settings, it was also similar in the other settings. Our result is similar to the results of the study conducted by Oliveira et al. (2020) (also analyzing the participants' flow experience in a gamified system). Both studies corroborate that different designs (personalization settings) can affect the flow experience dimensions differently; however, this difference is not always significant and it is not yet possible to identify exactly what can cause the effects.

One of the most exposed differences (visually in Fig. 1) was in the dimensions of the transformation of time and *autotelic* experience, where male participants using the default version had high experiences in comparison to the other participants.

Table 5 Post hoc tests

	Test statistic	Std. test statistic	<i>p</i> value	Adj. <i>p</i> value
FM-MD	- 57,849	- 3,163	0.002	0.023

FM female using the male-personalized version, *MD* male using the neutral version

**Fig. 4** Independent samples Kruskal–Wallis test

However, this difference also was not significant, that is, despite the visual difference, transformation of time, and *autotelic* experience were not affected by any of the personalization (see Table 4). Similar, different studies, for instance, the study conducted by Erhel and Jamet (2019) and the study conducted by Oliveira et al. (2020), most of the flow experience dimensions were not affected for different gamification designs.

The only dimension where results presented a significant difference was the action–awareness merging dimension, where females using the male-stereotyped version had their action–awareness merging negatively affected (see Tables 4, 5, Figs. 3, and 4). This result can be bought from the study conducted by Albuquerque et al. (2017), which found that male-stereotyped environments increased females’ anxiety. Initially, the action–awareness merging dimension, according to Jackson et al. (2011), represents a unit of complete awareness of the task being performed. Thus, it can be related to anxiety which, depending on the anxiety level, can result in a loss of consciousness in the task that is being performed.

In both studies [Albuquerque et al. (2017)’ study and our study], female participants using the male-stereotyped version had what we might call a negative experience (in the case of Albuquerque et al. (2017), with increased anxiety, in our case with action–awareness merging being negatively affected). This result draws attention to the fact that different experiences (*e.g.*, anxiety and action–awareness merging) of females when using male-stereotyped systems can be negatively affected. Such a result may have implications because in general systems that are not personalized tend to have characteristics more similar to the male audience (Wanner et al.,

2020; Albuquerque et al., 2017). So our result also draws attention to the importance of improving personalization systems for females.

Concerning users' performance, the results indicate that performance was not affected by the different stereotyped versions (see Table 4). This result may have occurred due to different factors. One of these factors is the fact that the users' performance can be linked to the intrinsic motivation (Bodkyn & Stevens, 2015; Chen & Law, 2016) to carry out the activities, that is, the users' worked to do the activities regardless of the system design. Another factor may be related to the type of activities performed by users in the study (quizzes on logical reasoning). That is, users may have been motivated only by the type of question being answered.

In recent studies, other researchers have found similar results, for example, both in the study of Oliveira et al. (2020) and in the study of Albuquerque et al. (2017), the different designs of the system did not affect the participants' performance. This result may indicate that participants' performance is not directly related to other experiences and that different studies need to be done to analyze how different types of personalization can affect users' performance in educational systems.

Of the nine flow experience dimensions, only one was affected by the stereotypes in the gamified educational system. Some general factors may be related to the results of the study. One of the factors is that in this study we chose to stereotype the system only in terms of colors and avatars, not personalized, for example, the gamification elements presented in the system. Recent studies have hypothesized that for system personalization to positively affect different types of users' experiences, personalization needs to be based on different factors [*e.g.*, gender, user types, and type of activities, among others (Hallifax et al., 2019a, b; Rodrigues et al., 2020)].

In summary, the results of our study demonstrate that stereotyped gamified educational systems affect the users' action–awareness merging and do not affect the flow experience and performance of users. These results highlight the importance of conducting new studies that can carry out similar experiments considering different types of personalization and analyzing different experiences.

Threats to validity and limitations

Because we conducted a study with humans, some threats inherent in the type of study were generated. To mitigate these threats as well as facilitate the reproduction of the study in different situations, we will describe below the threats to validity identified in this study. Initially, our study evaluated a subjective human experience (*i.e.*, flow experience) that can be difficult to measure. To mitigate this threat, we used only validated instruments widely used in scientific studies [*i.e.*, the FSS proposed by Jackson and Eklund (2002) and validated by Hamari and Koivisto (2014) for gamification domain].

In this study, we asked the participant's gender. In this sense, several genders could be available as an option, yet we used only two gender options (male and female), which may not have contemplated some participants. To mitigate this limitation, we enter the options "other" and "I prefer not to inform" for participants who do not feel represented in the male or female options. No participant marked any

option other than “male” or “female.” We intend to include other options in future studies, to obtain the possibility of new insights.

At the same time, we understand that gender in this context is superficially linked to sex and that numerous studies address the issue in a much deeper way, such as the work of De Lauretis (1987). However, as the first study in this line in the area of gamification applied to education, we believe that we are setting a precedent for future work to investigate the subject, which is so complex, in greater depth.

People from different countries answered the survey. These people may have different preferences about “what is a stereotyped system for males or females?” To mitigate this threat, we stereotyped the system according to more general aspects (e.g., colors and avatars) based on international studies, considering data from people from different countries. However, even so, we suggest comparative studies between people from different countries (cross-cultural studies) be done. Studies replicating the experiment conducted by Oyibo et al. (2017) in the field of education can help advance the state-of-the-art and understand the influence of culture on the effect of age and gender on educational technologies design.

Finally, the system gamification design was not planned based on systematic studies (e.g., gamification design frameworks) which can negatively influence the quality of the system and how the participants perceived the system. Likewise, the system’s personalization may have been limited concerning the real needs of the participants (in terms of stereotypes). To mitigate this threat, we used a system already used in other experimental studies that based most choices on other studies in the area.

Concerning the study limitations, the study was conducted by people of different ages and cultures, not allowing us to identify whether the results apply to people of specific ages or cultures. This limitation opens space for conducting new studies that replicate this experiment in controlled environments (e.g., with participants of the same age group in a specific school). Another limitation of the study is in the type of system and the time the study took place, as the system is for quick use (casual system), for example, not allowing the user to login and log out of the system several times and can use the system in different moments. This limitation opens space for the replication of the study in larger systems, which allow, for example, the conduct of longitudinal studies, where the participants’ flow experience can be measured at various times throughout the use of the system.

Building the future

Based on our study results and limitations, we identified different insights to drive future studies in this field. Thus, in this section, we present a research agenda to advance the research field.

- *Multiple factor-based personalization* In recent studies (Albuquerque et al., 2017; Oliveira et al., 2020; Erhel & Jamet, 2019), instruments used in the experiments were personalized/stereotyped based on specific factors (e.g., gamer types, gender, age). However, few studies have analyzed how different factors applied together (e.g., gender-based and age-based personalization) affect the users’

experience. Thus, *we suggest that future studies should analyze how multiple factors-based personalizations affect users' experience.*

- *Analysis of new experiences* As it is a relatively new area, with potential for growth, so far few experiences have been analyzed in the studies carried out (e.g., motivation, anxiety, and flow). However, several other experiences can be related to the learning experience and that can be evaluated in future studies. For this reason, *we recommend that future studies also focus on analyzing the effects of customizing gamified educational systems on other user experiences.*
- *Cross-cultural studies* Recently, studies have shown that personalization/stereotypes affects users' experience according to cultural variables (Oyibo et al., 2017; Toda et al., 2020; Oyibo & Vassileva, 2020). Concerning the effects of stereotyped gamified educational systems on users' experience according to cultural factors, results are still incipient. Thus, *we suggest that future studies analyze how personalized gamified educational systems affect users' experience according to cultural factors.*
- *Longitudinal studies* Most of the studies conducted so far are studies carried out in a short period. With the advances in the area, we consider it important that future studies can analyze the effects of gender-stereotyped educational systems within long-term interventions. Therefore, *we suggest that longitudinal studies can be conducted in future.*

Concluding remarks

In recent years, different studies have been conducted to analyze whether gender stereotype-based interfaces affect users' experience in educational systems. However, there is still no concrete answer to this question and many studies have highlighted the challenge of understanding when the stereotyped gamified educational systems affect users' flow experience and performance. Thus, in this study, we analyzed the effects of gender stereotype-based interfaces on users' flow experience and performance in a gamified educational system.

Answering our first research question (Do gender stereotype-based interfaces affect the users' flow experience in a gamified educational system?), we identified that gender-stereotyped personalization in gamified educational systems affects users' action–awareness merging, however, does not affect users' overall flow experience. Answering our second research question (Do gender stereotype-based interfaces affect the users' performance in gamified educational systems?), the results indicate that the gender-stereotyped gamified educational systems do not affect users' performance.

From the results obtained, we intend, in future works to perform more in-depth studies involving the design of interfaces based on the most current theories of gender, which go beyond sex. This vision can advance the literature on gamification applied to education and toward an equitable design. We also aim to replicate the study by considering other aspects of personalization (e.g., considering other gamification elements and content-based gamification), as well as evaluating other aspects of user experience. Finally, we will replicate the experiment through a longitudinal

study, get the users’ flow experience at different times, and analyze it using different machine learning techniques.

Appendix 1: System examples

Examples of the different versions of the system used in our study (Fig. 5).

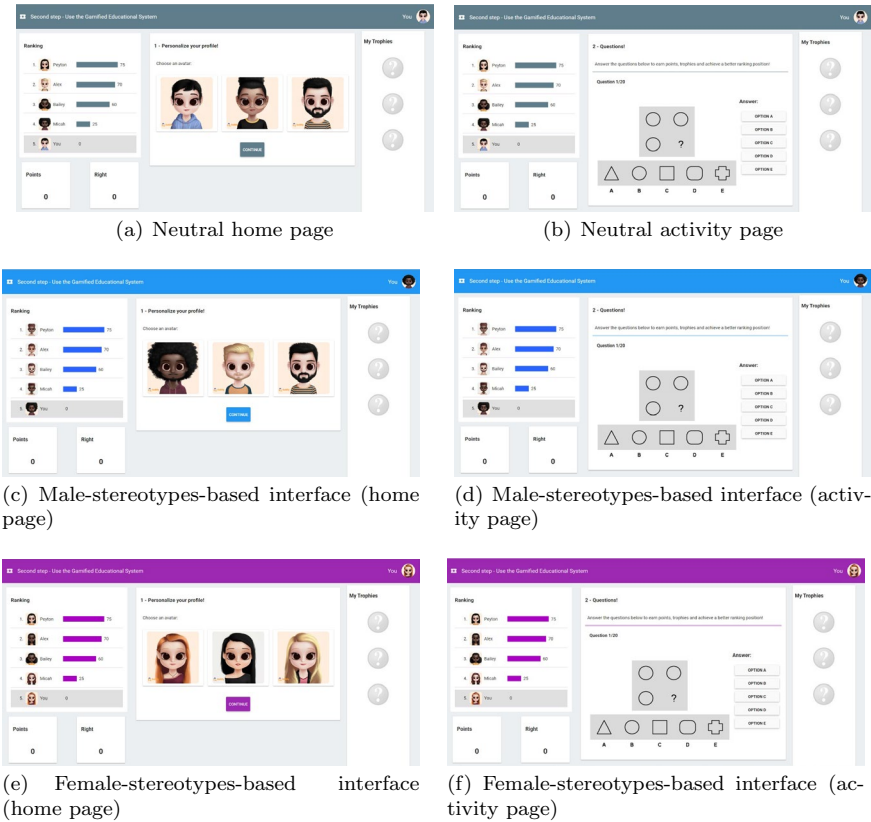


Fig. 5 System example

Appendix 2: Flow state scale

The file containing the FSS [developed by Jackson and Eklund (2002) and validated by Hamari and Koivisto (2014)] was used in our study.

Please, answer the questions below by selecting the alternatives that best fit your psychological state during the reactivation of activities in the educational system. Consider the scale from 1 to 5, respectively, Never, Rarely, Sometimes, Frequently, and Always as given in Table 6.

Table 6 Flow state scale

Dimensions	Questions	1	2	3	4	5
CSB	I felt I was competent enough to meet the demands of the situation					
MMA	I did things spontaneously and automatically without having to think					
G	I had a strong sense of what I wanted to do					
F	I had a good idea about how well I was doing while I was involved in the task/activity					
C	I was completely focused on the task at hand					
CTRL	I had a feeling of total control over what I was doing					
LSC	I was not worried about what others may have been thinking of me					
T	The way time passed seemed to be different from normal					
A	I found the experience extremely rewarding					
	If you are filling out the form carefully, answer 4					

*Validation question (presented in a random order); *CSB* challenge–skill balance, *MMA* action–awareness merging, *G* clear goals, *F* unambiguous feedback, *C* total concentration on the task at hand, *CTRL* sense of control, *LSC* loss of self-consciousness, *T* transformation of time, *A* autotelic experience

Author contributions WO contributed to conceptualization, methodology, formal analysis, validation, and writing, reviewing, & editing of the manuscript. JH contributed to supervision, validation, formal analysis, and writing, reviewing, & editing of the manuscript. WF contributed to system design and formal analysis. AMT contributed to writing, reviewing, and editing of the manuscript. PTP contributed to writing, reviewing, and editing of the manuscript. JV contributed to supervision and writing, reviewing, and editing of the manuscript. SI contributed to supervision and writing, reviewing, and editing of the manuscript.

Funding São Paulo Research Foundation (FAPESP Projects: 2018/07688-1 and 2020/02801-4), Academy of Finland Flagship Programme [Grant No. 337653 - Forest-Human-Machine Interplay (UNITE)], CNPq (Project: 308458/2020-6).

Data availability Original dataset available as supplementary material.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no conflicts of interest to disclose.

Ethical statement I hereby declare that this research paper is the authors' own original work, which has not been previously published elsewhere. The research paper is not currently being considered for publication elsewhere. The paper reflects the authors' own research and analysis in a truthful and complete manner. The responsibility of this statement shall be borne by the authors of the paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Albuquerque, J., Bittencourt, I. I., Coelho, J. A., & Silva, A. P. (2017). Does gender stereotype threat in gamified educational environments cause anxiety? An experimental study. *Computers & Education*, *115*, 161–170.
- Alhadreti, O., & Mayhew, P. (2016). "Thinking about thinking aloud": An investigation of think-aloud methods in usability testing. In *Proceedings of the 30th International BCS Human Computer Interaction Conference* (Vol. 30, pp. 1–3).
- Aljedaibi, W., & Khamis, S. (2019). Towards measuring the project management process during large scale software system implementation phase. *ISeCure*, *11*(3), 161–172.
- Bai, S., Hew, K. F., & Huang, B. (2020). Is gamification "bullshit"? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, 100322.
- Basow, S. A. (1992). *Gender: Stereotypes and roles*. Thomson Brooks/Cole Publishing Co.
- Beigpour, S., & Pedersen, M. (2015). Color play: Gamification for color vision study. In *Proceedings of Color and Image, Midterm Meeting of the International Colour Association* (pp. 1–6).
- Belk, M., Germanakos, P., Fidas, C., & Samaras, G. (2014). A personalization method based on human factors for improving usability of user authentication tasks. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 13–24). Springer.
- Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*(1), 78–117.
- Bodkyn, C., & Stevens, F. (2015). Self-directed learning, intrinsic motivation and student performance. *The Caribbean Teaching Scholar*, *5*(2).
- Buil, I., Catalán, S., & Martínez, E. (2019). The influence of flow on learning outcomes: An empirical study on the use of clickers. *British Journal of Educational Technology*, *50*(1), 428–439.
- Caldiera, V. R. B. G., & Rombach, H. D. (1994). Goal question metric paradigm. *Encyclopedia of Software Engineering*, *1*, 528–532.
- Carrasco, R., Baker, S., Waycott, J., & Vetere, F. (2017). Negotiating stereotypes of older adults through avatars. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (pp. 218–227).
- Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education: A Journal of Educational Research and Practice*, *12*(2).
- Chen, C. H., & Law, V. (2016). Scaffolding individual and collaborative game-based learning in learning performance and intrinsic motivation. *Computers in Human Behavior*, *55*, 1201–1212.
- Coursaris, C. K., Swierenga, S. J., & Watrall, E. (2008). An empirical investigation of color temperature and gender effects on web aesthetics. *Journal of Usability Studies*, *3*(3), 103–117.

- Csikszentmihalyi, M. (1997a). *Finding flow: The psychology of engagement with everyday life*. Basic Books.
- Csikszentmihalyi, M. (1997b). Flow and Education. *NAMTA Journal*, 22(2), 2–35.
- Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety*. Jossey-Bass.
- Csikszentmihalyi, M. (2014a). *Applications of flow in human development and education*. Springer.
- Csikszentmihalyi, M. (2014b). Toward a psychology of optimal experience. In *Flow and the foundations of positive psychology* (pp. 209–226). Springer.
- Csikszentmihalyi, M., & Csikszentmihalyi, I. (1975). *Beyond boredom and anxiety* (Vol. 721). Jossey-Bass.
- De Lauretis, T. (1987). *Technologies of gender: Essays on theory, film, and fiction*. Indiana University Press.
- Deaux, K., & Lewis, L. L. (1984). Structure of gender stereotypes: Interrelationships among components and gender label. *Journal of Personality and Social Psychology*, 46(5), 991.
- Denden, M., Tlili, A., Essalmi, F., Jemni, M., Chen, N. S., & Burgos, D. (2021). Effects of gender and personality differences on students' perception of game design elements in educational gamification. *International Journal of Human-Computer Studies*, 154, 102674.
- Dorji, U., Panjaburee, P., & Srisawasdi, N. (2015). Gender differences in students' learning achievements and awareness through residence energy saving game-based inquiry playing. *Journal of Computers in Education*, 2(2), 227–243.
- Erhel, S., & Jamet, E. (2019). Improving instructions in educational computer games: Exploring the relations between goal specificity, flow experience and learning outcomes. *Computers in Human Behavior*, 91, 106–114.
- Gao, B., Wan, Q., Chang, T., & Huang, R. (2019). A framework of learning activity design for flow experience in smart learning environment. In *Foundations and trends in smart learning* (pp. 5–14). Springer.
- Göbl, B., Comber, O., Felber, J., Fenkart, H., Mayer, H., & Motschnig, R. (2021). Dig-equality ff-a playful approach for researching and fostering gender education in secondary schools. In *2021 IEEE Global Engineering Education Conference (EDUCON)* (pp. 964–967). IEEE.
- Greene, K. S., & Gynther, M. D. (1995). Blue versus periwinkle: Color identification and gender. *Perceptual and Motor Skills*, 80(1), 27–32.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L., et al. (1998). *Multivariate data analysis* (Vol. 5). Prentice Hall.
- Hallifax, S., Serna, A., Marty, J. C., & Lavoué, É. (2019a). Adaptive gamification in education: A literature review of current trends and developments. In *European Conference on Technology Enhanced Learning* (pp. 294–307). Springer.
- Hallifax, S., Serna, A., Marty, J. C., Lavoué, G., & Lavoué, E. (2019b). Factors to consider for tailored gamification. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (pp. 559–572). ACM.
- Hallock, J. (2003). *Colour assignment-associations*. Colour Assignment.
- Hamari, J. (2019). *Gamification, The Blackwell Encyclopedia of Sociology, chap Gamification* (pp. 1–3). <https://doi.org/10.1002/9781405165518.wbeos1321>.
- Hamari, J., & Koivisto, J. (2014). Measuring flow in gamification: Dispositional flow scale-2. *Computers in Human Behavior*, 40, 133–143.
- Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on System Sciences* (pp. 3025–3034). IEEE.
- Hassan, L., Jylhä, H., Sjöblom, M., & Hamari, J. (2020). Flow in vr: A study on the relationships between preconditions, experience and continued use. In *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 1196–1205).
- Heutte, J., Fenouillet, F., Martin-Krumm, C., Boniwell, I., & Csikszentmihalyi, M. (2016). Proposal for a conceptual evolution of the flow in education (eduflow) model. In *8th European Conference on Positive Psychology (ECPP 2016)* (pp. 1–34).
- Hsieh, Y. H., Lin, Y. C., & Hou, H. T. (2016). Exploring the role of flow experience, learning performance and potential behavior clusters in elementary students' game-based learning. *Interactive Learning Environments*, 24(1), 178–193.
- Hurlbert, A. C., & Ling, Y. (2007). Biological components of sex differences in color preference. *Current Biology*, 17(16), R623–R625.
- Jackson, S., Eklund, B., & Martin, A. (2011). *The flow manual the manual for the flow scales manual* (pp. 1–85). Mind Garden.

- Jackson, S. A., & Eklund, R. C. (2002). Assessing flow in physical activity: The flow state scale-2 and dispositional flow scale-2. *Journal of Sport and Exercise Psychology*, 24(2), 133–150.
- Jackson, S. A., & Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: The flow state scale. *Journal of Sport and Exercise Psychology*, 18(1), 17–35.
- Janelli, M., & Lipnevich, A. A. (2021). Effects of pre-tests and feedback on performance outcomes and persistence in massive open online courses. *Computers & Education*, 161, 104076.
- Karniol, R. (2011). The color of children's gender stereotypes. *Sex Roles*, 65(1–2), 119–132.
- Khoshkangini, R., Marconi, A., & Valetto, G. (2017). Machine learning for personalized challenges in a gamified sustainable mobility scenario. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play* (pp. 361–368). ACM.
- Kiili, K., de Freitas, S., Arnab, S., & Lainema, T. (2012). The design principles for flow experience in educational games. *Procedia Computer Science*, 15, 78–91.
- Klock, A. C. T., Gasparini, I., Pimenta, M. S., & Hamari, J. (2020). Tailored gamification: A review of literature. *International Journal of Human-Computer Studies*, 102495.
- Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, 45, 191–210.
- Kollmayer, M., Schober, B., & Spiel, C. (2018). Gender stereotypes in education: Development, consequences, and interventions. *European Journal of Developmental Psychology*, 15(4), 361–377.
- Komalawardhana, N., & Panjaburee, P. (2018). Proposal of personalised mobile game from inquiry-based learning activities perspective: Relationships among genders, learning styles, perceptions, and learning interest. *International Journal of Mobile Learning and Organisation*, 12(1), 55–76.
- Lavoué, É., Monterrat, B., Desmarais, M., & George, S. (2018). Adaptive gamification for learning environments. *IEEE Transactions on Learning Technologies*, 12(1), 16–28.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*.
- Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis*. Lawrence Erlbaum Associates Publishers.
- Masthoff, J., & Vassileva, J. (2015). Tutorial on personalization for behaviour change. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (pp. 439–442). ACM.
- Nah, F. F. H., Zeng, Q., Telaprolu, V. R., Ayyappa, A. P., & Eschenbrenner, B. (2014). Gamification of education: a review of literature. In *International conference on HCI in business* (pp. 401–409). Springer.
- Oliveira, W., & Bittencourt, I. I. (2019). *Tailored gamification to educational technologies*. Springer.
- Oliveira, W., Bittencourt, I. I., Isotani, S., Dermeval, D., Marques, L. B., & Silveira, I. F. (2018). Flow theory to promote learning in educational systems: Is it really relevant? *Brazilian Journal of Computers in Education*, 26(02), 29.
- Oliveira, W., Pastushenko, O., Rodrigues, L., Toda, A. M., Palomino, P. T., Hamari, J., & Isotani, S. (2021). Does gamification affect flow experience? a systematic literature review. In *GamiFIN Conference 2021: Proceedings of the 5th International GamiFIN Conference* (pp. 1–10).
- Oliveira, W., Toda, A., Palomino, P., Rodrigues, L., Isotani, S., & Shi, L. (2019). Towards automatic flow experience identification in educational systems: A theory-driven approach. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (pp. 581–588). ACM.
- Oliveira, W., Toda, A., Palomino, P., Shi, L., Isotani, S., Bittencourt, I. I., & Vassileva, J. (2020). Does tailoring gamified educational systems matter? The impact on students' flow experience. *Hawaii International Conference on System Sciences*, 20, 1226–1235.
- Orji, R. (2014). *Design for behaviour change: A model-driven approach for tailoring persuasive technologies*. PhD thesis, University of Saskatchewan
- Orji, R., Mandryk, R. L., & Vassileva, J. (2014). Selecting effective strategies for tailoring persuasive health games to gamer types. *Graphics, Animation and New Media*.
- Orji, R., Mandryk, R. L., Vassileva, J., & Gerling, K. M. (2013). Tailoring persuasive health games to gamer type. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2467–2476). ACM.
- Oyibo, K., Orji, R., & Vassileva, J. (2017). The influence of culture in the effect of age and gender on social influence in persuasive technology. In *Adjunct Publication of the 25th Conference on User Modeling* (pp. 47–52). Adaptation and Personalization: ACM.
- Oyibo, K., & Vassileva, J. (2020). The effect of layout and colour temperature on the perception of tourism websites for mobile devices. *Multimodal Technologies and Interaction*, 4(1), 8.
- Palan, S., & Schitter, C. (2018). Prolific. ac-a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27.

- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5), 411–419.
- Rapp, A., Hopfgartner, F., Hamari, J., Linehan, C., & Cena, F. (2019). Strengthening gamification studies: Current trends and future opportunities of gamification research. *International Journal of Human-Computer Studies*, 127, 1–6.
- Rodrigues, L., Toda, A. M., Palomino, P. T., Oliveira, W., & Isotani, S. (2020). Personalized gamification: A literature review of outcomes, experiments, and approaches. In *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 699–706).
- Rodríguez-Ardura, I., Meseguer-Artola, A. (2019). Flow experiences in personalised e-learning environments and the role of gender and academic performance. *Interactive Learning Environments*, 1–24.
- Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, 69, 371–380.
- Santos, J., Andrade, E., Benevides, K., Silva, K., Nascimento, J., Bittencourt, I., Pereira, M., Fernandes, S., & Isotani, S. (2022). Does gender stereotype threat affects the levels of aggressiveness, learning and flow in gamified learning environments? An experimental study. *Education and Information Technologies*, 1–26.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611.
- 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). ACM.
- Steele, C. M. (2011). *Whistling Vivaldi: How stereotypes affect us and what we can do*. WW Norton & Company.
- Toda, A., Klock, A.C.T., Palomino, P. T., Rodrigues, L., Oliveira, W., Stewart, C., Cristea, A. I., Gasparini, I., & Isotani, S. (2020). Gamicsm: Relating education, culture and gamification—a link between worlds. In *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems* (pp. 1–10).
- Toda, A. M., Klock, A. C., Oliveira, W., Palomino, P. T., Rodrigues, L., Shi, L., Bittencourt, I., Gasparini, I., Isotani, S., & Cristea, A. I. (2019). Analysing gamification elements in educational environments using an existing gamification taxonomy. *Smart Learning Environments*, 6(1), 16.
- Toda, A. M., Oliveira, W., Shi, L., Bittencourt, I. I., Isotani, S., & Cristea, A. (2019). Planning gamification strategies based on user characteristics and dm: A gender-based case study. *Proceedings of the Educational Data Mining, 2019*, 438–443.
- Tondello, G. F., Mora, A., & Nacke, L. E. (2017). Elements of gameful design emerging from user preferences. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (pp. 129–142). ACM.
- Tuah, N. M., & Wills, G. B. (2020). Measuring the application of anthropomorphic gamification for transitional care; a goal-question-metric approach. In *Computational Science and Technology* (pp. 553–564). Springer.
- Tuunanen, J., & Hamari, J. (2012). Meta-synthesis of player typologies. In *Proceedings of Nordic Digra 2012 Conference: Games in Culture and Society* (pp. 1–14).
- Vail, A. K., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2015). The mars and venus effect: The influence of user gender on the effectiveness of adaptive task support. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 265–276). Springer.
- Van Solingen, R., Basili, V., Caldiera, G., & Rombach, H. D. (2002). Goal question metric (gqm) approach. *Encyclopedia of Software Engineering*.
- Vidergor, H. E. (2021). Effects of digital escape room on gameful experience, collaboration, and motivation of elementary school students. *Computers & Education*, 166, 104156.
- Wanner, T., Wanner, T., & Etzold, V. (2020). Effects on girls' emotions during gamification tasks with male priming in stem subjects via eye tracking. In *Smart Education and e-Learning 2020* (pp. 67–78). Springer.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in Software Engineering*. Springer.
- Yoshida, K., Asakawa, K., Yamauchi, T., Sakuraba, S., Sawamura, D., Murakami, Y., & Sakai, S. (2013). The flow state scale for occupational tasks: Development, reliability, and validity. *Hong Kong Journal of Occupational Therapy*, 23(2), 54–61.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Dr. Wilk Oliveira is a researcher of the Gamification Group at the Faculty of Information Technology and Communication Sciences at Tampere University. He worked on a series of research projects, maintained by important international funding agencies, in which he has been working on the development, application, and evaluation of various educational technologies, producing different products and scientific publications in some of the most important international conferences and journals in the field of educational technologies. His productions have generated several academic merit awards, including the best paper in international conferences and presentations for international study programs. In the public sector, he collaborated with the Brazilian Ministry of Education on projects related to the development, application, and evaluation of educational technologies. He is also an enthusiast and precursor of the Virtualization of Educational Games in Brazil, becoming the co-founder of the first Brazilian start-up in the segment, called Virtualize Games. Recently, he created the start-up Eagle-Edu, the first evidence-based gamification edutech. Wilk believes in education to improve the world, seeking through his research the improvement of the quality and access to education. His main areas of expertise are: Educational Technologies (specifically, but not exclusively, Flow Theory, Gamification and Educational Games) and Computing Education.

Juho Hamari is a professor of gamification and the leader of the Gamification Group at the Faculty of Information Technology and Communication Sciences at Tampere University. Professor Hamari is leading research on gamification under the Academy of Finland Flagship, Profiling, and Center of Excellence programs. His and his group's research cover several forms of information technologies, such as games, motivational information systems, new media (social networking services, eSports), peer-to-peer economies (sharing economy, crowdsourcing), and virtual economies. Dr. Hamari has authored several seminal empirical, theoretical, and meta-analytical scholarly articles on these topics from perspective of consumer behavior, human–computer interaction, game studies, and information systems science.

William Ferreira Computer Engineer with an emphasis in Data Science from the University of São Paulo. Passionate about projects that add value to people's lives. He is currently a data engineer at Bradesco Bank (Brazil), where he works as an Engineer in Machine Learning.

Armando M. Toda is a PDRA at Durham University, and research collaborator at University of São Paulo (USP). Has a B.Sc., M.Sc., and Ph.D in Computer Science, the latest from the University of São Paulo. His research career is invested in the design and application of gamification in learning environments. His scientific and social mission is to demonstrate how game-based technologies can improve people's life and how these technologies can impact positively in education. Has published many scientific papers in journals and conferences, which won international recognition and awards. His research interests are aligned but not limited to, gamification, human–computer interaction, learning environments, artificial intelligence applied to education, educational data mining, and computational thinking.

Dr. Paula Palomino is a post-doctorate researcher at Computing Institute of Federal University of Alagoas (Brazil) and a visiting researcher at Stratford School of Interaction Design and Business of University of Waterloo (Canada), holding a Ph.D. in Computer Science at University of São Paulo. Also holds an M.Sc and B.Sc in Communication. Her research interests are game design, gamification, UX and digital culture, specifically researching the use of subjective game elements in gamification frameworks applied to educational contexts.

Julita Vassileva Julita Vassileva is a Professor in Computer Science at the University of Saskatchewan, Canada, and a leading researcher in AI in education, user modeling and personalization, social computing, persuasive technology and behavior change. She has authored over 300 research papers and is the founding Editor-in-Chief of Frontiers in AI section on AI Supporting Human Learning and Behavior Change. Her group currently explores methods for increasing student engagement in e-learning by modeling motivation and engagement by mining learning analytics data, designing gamified and game-based learning systems and deploying personalized socially-oriented persuasive strategies.

Seiji Isotani is a Visiting Professor of Education at the Harvard Graduate School of Education and a full Professor of Computer Science and Learning Technology at the University of Sao Paulo, Brazil. He received B.Sc. and M.Sc. degrees in Computer Science from the University of Sao Paulo, Brazil and a Ph.D. degree in Information Engineering from Osaka University, Japan. His research career has been devoted to the conception, design, development, testing, and deployment of intelligent and collaborative educational systems using ontologies and other semantic technologies. His scientific and social missions converge into a single objective: to enable the realization of Anytime, Anywhere, Anybody Learning (AAAL) by developing cutting-edge technology. He is the co-founder of two startups, which have won several innovation awards in the fields of education and the semantic web. He has published over a hundred scientific papers, books, and book chapters on educational technology and the semantic web. His main research interests are in the areas of gamification, ontological engineering, computer-supported collaborative learning (CSCL), artificial intelligence in education (AIED), and technology-enhanced learning.