Computers & Education 78 (2014) 66–77

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance

Katheryn R. Christy*, Jesse Fox [1]

School of Communication, The Ohio State University, 3016 Derby Hall, 154 N Oval Mall, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 31 December 2013
Received in revised form
12 May 2014
Accepted 15 May 2014
Available online 24 May 2014

Keywords:
Distance education and telelearning
Gender studies
Human–computer interface
Interactive learning environments
Teaching/learning strategies

ABSTRACT

Gamification includes the use of gaming features, such as points or leaderboards, in non-gaming contexts, and is a frequently-discussed trend in education. One way of gamifying the classroom is to introduce leaderboards. Leaderboards allow students to see how they are performing relative to others in the same class. Little empirical research has investigated the impact of leaderboards on academic performance. In this study, 80 female undergraduates took a math test in a virtual representation of a classroom after being exposed to one of three leaderboard conditions: a leaderboard where men held the majority of the top positions, a leaderboard where women held the majority of top positions, and a no leaderboard condition. Participants in the female majority leaderboard condition performed more poorly on the math test than those in the male leaderboard condition, yet demonstrated a higher level of academic identification than those in the male and control conditions. The authors conclude with a discussion of the implications that this study's findings may have for the use of leaderboards within educational environments.

1. Introduction

Gamification is the "use of game design elements in non-game contexts" (Deterding, Dixon, Khaled, & Nacke, 2011, n. p.), and many have discussed the benefits of gamifying the classroom (Connolly, Boyle, MacArthur, Haines, & Boyle, 2012; Deterding et al., 2011; Kapp, 2012; Landers & Callan, 2011; Lee & Hammer, 2011; Muntean, 2011). Supporters of gamification claim that the use of game elements in the classroom enhances learning by increasing engagement and motivation and facilitating social learning (Muntean, 2011), and encourage teachers to integrate these methods into their classrooms (Jones, 2010; Kaya, 2010; Salter, 2011a, 2011b, 2011c). Emerging studies, however, suggest that gamifying learning may not always be beneficial (e.g., de-Marcos, Domínguez, Saenz-de-Navarrete, Pagés, 2014; Domínguez et al., 2013; for a review, see Hamari, Koivisto, & Sarsa, 2014).

One way of gamifying the classroom is to introduce leaderboards to the learning environment (Hamari et al., 2014). Leaderboards allow students to see how they are performing relative to others in the same class; some argue that the competitive environment this creates is beneficial to learning (Camilleri, Busuttil, & Montebello, 2011; Charsky, 2010; Kapp, 2012; Muntean, 2011). However, little empirical research has investigated the impact of leaderboards on academic performance or the consequences arising from leaderboard-prompted comparisons. Indeed, it is possible that the use of leaderboards within educational settings may create high levels of stereotype threat or detrimental upward social comparisons. Research has demonstrated that both stereotype threat (Nguyen & Ryan, 2008; Sekaquaptewa & Thompson, 2003; Spencer, Steele, & Quinn, 1999; Thompson & Sekaquaptewa, 2002) and upward social comparisons (Dijkstra, Kuyper, van der Werf, Buunk, & van der Zee, 2008; Lyubomirsky, Kasri, & Zehm, 2003; Muller & Fayant, 2010) can have a detrimental influence on students' academic performance. This study aims to examine whether or not interaction with a leaderboard produces effects consistent with stereotype threat or social comparison experiences.

* Corresponding author. Tel.: +1 724 889 5101.
E-mail addresses: christy.89@osu.edu, katheryn.christy@gmail.com (K.R. Christy), fox.775@osu.edu (J. Fox).
[1] Tel.: +1 614 292 2055.

http://dx.doi.org/10.1016/j.compedu.2014.05.005
0360-1315/

2. Gamification, education, and leaderboards

Gamification has become a commonly recommended pedagogical tool (Anderson & Rainie, 2012; Boulet, 2012; Chou, 2013; Kapp, 2012), and advice on how to implement it in the classroom regularly appears in *The Chronicle of Higher Education*, a publication geared toward post-secondary school instructors (e.g., Jones, 2010; Kaya, 2010; Salter, 2011a, 2011b, 2011c). At its heart, gamification focuses on the idea of taking some of the elements that make games engaging (e.g., direct competition, visible rewards) and applying them to non-game contexts (Deterding et al., 2011). Gamification is theorized to increase student motivation by providing students with clear, achievable goals (Landers & Callan, 2011), by making learning environments more fun and engaging (Bajdor & Dragolea, 2011; Cohen, 2011; Landers & Callan, 2011; Muntean, 2011), and by encouraging competition (Hamari, 2013; Reeves & Read, 2009).

However, recent research has suggested that gamification may not always be an effective tool. Several studies have suggested that the positive results found in gamification studies may have been the result of novelty effects, as the impact of gamification seems to taper off as time goes on (Farzan et al., 2008; Hamari, 2013; Koivisto & Hamari, 2014). Other studies have found that gamification actually decreases class participation, and results in poorer performance on exams (Domínguez et al., 2013; de-Marcos et al., 2014). In this light, further research on the impact of gamification on learning outcomes is warranted.

One of the most popular gamification techniques is including a leaderboard in the learning environment (Hamari et al., 2014). A leaderboard is a game design element consisting of a visual display that ranks players according to their accomplishments; when used in an educational setting it serves as a way for students to directly compare their own performance with that of others (for an example, see Fig. 1). Indeed, past research has shown that leaderboards increase competition (Butler, 2013) and stimulate social comparisons (Costa, Wehbe, Robb, & Nacke, 2013) when used in a business context. Although these comparisons may be beneficial in business contexts, they may result in unintended consequences in academic environments. More specifically, the use of leaderboards within educational settings may act to create high levels of stereotype threat or present opportunities for upwards social comparison, either of which could have a negative

TOP TEN LEADERBOARD	
Username	**Score**
1. JessicaP	20
2. KateRC	20
3. RachelE	19.5
4. TonyC	19.5
5. AlexisN	19.5
6. MeganAB	19
7. LauraMC	19
8. KateAM	19
9. MikeRH	18
10. TedB	18

Fig. 1. An example leaderboard (here, female-dominated).

impact on students' academic performance. Although previous research has demonstrated that both stereotype threat and upward social comparisons can have a negative effect on academic performance, we could identify no studies that investigated the possibility that leaderboards may trigger these reactions.

3. Theoretical explanations

3.1. Stereotype threat

Stereotype threat is a phenomenon that occurs when "one can be judged by, treated in terms of, or self-fulfill negative stereotypes about one's group" (Spencer et al., 1999, p. 6). Being under stereotype threat tends to result in underperformance on the stereotyped domain (Spencer et al., 1999). Indeed, some research has found that the cause of this underperformance is that being under stereotype threat significantly decreases working memory capacity, which in turn negatively influences performance on cognitively taxing tasks (Beilock, Rydell, & McConnell, 2007; Johns, Inzlicht, & Schmader, 2008; Schmader & Johns, 2003).

The impact of stereotype threat on academic performance has been well-established, with one of the best understood situations being the influence of stereotype threat on women's math performance. Women routinely underperform men on standardized math tests and in math classrooms (College Board, 2012; Ellison & Swanson, 2010; Hedges & Nowell, 1995; Xie & Shauman, 2003). Research suggests that this occurs, at least in part, due to stereotype threat, the relevant stereotype being "women are worse at math than men" (see Nguyen & Ryan's, 2008 meta-analysis). Indeed, when women are informed that a math test is "gender fair" (i.e., that men and women both perform equally well on the test), the stereotype threat effect tends to dissipate, and women perform equally well as men (Sekaquaptewa & Thompson, 2003; Spencer et al., 1999). For example, in Spencer et al.'s (1999) study, participants were told that they were taking a math exam in order to investigate potential gender differences in math abilities among those who are good at math. Half the participants were then told that the test had demonstrated gender differences in the past (high stereotype threat), while the second half of the participants were told that the test had not shown gender differences in the past (low stereotype threat). Women performed significantly worse than men when informed that they were taking a test that showed gender differences, but performed equally well as men when told they were taking a test that did not show gender differences. A similar pattern is found for African American students; those who experience stereotype threat (i.e., "African Americans are less intelligent than Caucasians") tend to perform more poorly than their Caucasian classmates across a range of academic domains (Thompson & Sekaquaptewa, 2002). These effects have been found in both naturalistic and experimental situations (see Huguet & Régner, 2007).

Not only can stereotype threat directly impact academic performance, but it can also influence the degree to which a student identifies with academics in general or one subject in particular (Steele, 1997; Steele, Spencer, & Aronson, 2002). Identification with a domain is the degree to which one regards a domain to be an important part of their identity (Steele, 1997). Some scholars have argued that in order to manage the self-threat that stereotype threat produces, individuals may choose to disidentify with the stereotyped domain (e.g., "Math isn't really important anyway"; Steele, 1997; Steele et al., 2002). Disidentification can occur as both an acute (referred to as domain disengagement) and a chronic (referred to as domain disidentification) response to stereotype threat (Pronin, Steele, & Ross, 2004; Schmader, Johns, & Barquissau, 2004; Schmader, Major, & Gramzow, 2001; Steele et al., 2002). For example, a study by Woodcock, Hernandez, Estrada, and Schultz (2012) examined the relationship between stereotype threat and science identification among African American and Hispanic/Lantino/a students. Their study spanned 3 years and used a national longitudinal panel of minority science students. They found that the experience of stereotype threat (i.e., perceptions that Black and Hispanic individuals are less intelligent) predicted scientific disidentification, which itself predicted a lowered intention to go into a science-related career (Woodcock et al., 2012). Similar results were seen in Verkuyten and Thijs' (2004) study of ethnic Turkish students and Cundiff, Vescio, Loken, and Lo's (2013) study of female science majors, although these studies were cross-sectional (not longitudinal) in nature.

Due to stereotype threat's potential to significantly inhibit academic performance and affect identification with a variety of academic domains, any pedagogical practice that may increase the likelihood of students experiencing stereotype threat is one that should be carefully examined. Given that leaderboards provide indicators of others' success (and perhaps indications of others' failure by a lack of representation) it is possible that they may promote stereotype threat.

3.1.1. Stereotype threat and leaderboards

The primary way that a leaderboard is likely to contribute to the experience of stereotype threat is by making the relevant stereotype highly salient (Aaronson et al., 1999; Keller, 2002; Schmader et al., 2004; Spencer et al., 1999). Past experiments on stereotype threat have shown that providing participants with situational cues that make them more aware of the relevant stereotype increases the effects of stereotype threat (e.g., Aaronson et al., 1999; Keller, 2002). These situational cues can be very subtle. For example, Inzlicht and Ben-Zeev (2000) had women complete a math test in a 3-person group that either consisted of the participant and two other women or the participant and two men. Although no explicit information about women's math performance was provided, the mixed-sex condition was designed to increase the salience of the woman's female identity and the salience of stereotypes related to their group. As predicted, the women in the mixed-sex condition performed significantly worse on the test than the women in the same-sex condition.

In the case of leaderboards, the position of the names on a leaderboard may provide a situational cue that visually illustrates a relevant stereotype (e.g., the leaderboard in a math class is dominated by men), and thus triggers stereotype threat. This situation is especially problematic in environments where the stereotyped group member is in the minority within the environment. Additionally, the very experience of being under stereotype threat can cause disidentification in the threatened domain (Cundiff et al., 2013; Verkuyten & Thijs, 2004). Thus, if leaderboards cause students to experience stereotype threat, we would expect to see a drop in academic identification or subject-specific identification (e.g., math) due to that experience.

Conversely, it is possible that if a leaderboard *refutes* a stereotype (e.g., the top positions on the math class leaderboard are all held by women), then stereotype threat may actually *decrease*. Studies suggest that when individuals from a threatened group are exposed to a positive in-group role model immediately prior to completing a stereotype-relevant task, they perform significantly better than in the absence of such a role model (Marx & Roman, 2002; McIntyre, Paulson, & Lord, 2003; Taylor, Lord, McIntyre, & Paulson, 2011). For example, a study on stereotype threat carried out by Marx and Roman (2002) had either a female or male experimenter mention that they were the one who had created the difficult math test participants were about to take. Exposure to the male experimenter showed the expected stereotype threat effect (i.e., women underperformed men); however, when exposed to the female experimenter, women performed as equally well as men. Therefore, it is not unreasonable to assume that seeing a leaderboard where members of the stereotyped group hold the majority of top positions may actually diminish the influence of stereotype threat and result in better academic performance.

The potential for leaderboards to generate stereotype threat is not the only drawback to the use of leaderboards. Because they make others' performance visible, leaderboards could also encourage detrimental social comparisons.

3.2. Social comparison

Humans often compare themselves to other people when judging their own performance (Hoorens & Van Damme, 2012). Social comparison theory suggests that humans are driven to self-evaluate, and thus individuals compare themselves to others to reduce uncertainty and make accurate self-assessments (Festinger, 1954). Furthermore, rather than focusing comparative processes on equivalent others, people often engage in *downward social comparison* with others who are worse off or *upward social comparison* with others who are better off (Buunk & Gibbons, 2007). Downward comparisons can lead to positive affect and a sense of superiority given the target's lower status, whereas upward comparisons can have the opposite effect, evoking negative affect and self-concept as the individual comes to terms with his or her lower status in relation to the target (Buunk & Gibbons, 2007).

Social comparison is common in educational settings. The classroom naturally facilitates comparison by providing frequent evaluation and constant exposure to peers' performance and abilities (Wehrens et al., 2010). In elementary school, for example, students are often visibly rewarded for good behavior via sticker charts or other visible rewards. College students often inquire about exam score means to interpret their performance. Dijkstra et al. (2008) reviewed studies on social comparison in the classroom and found that students tended to identify students similar to themselves, especially along the dimensions of sex and age, and engage in upward social comparison, which then led to negative affect and lower academic identification. Other research suggests that upward comparisons can also result in ruminative thought and attentional disturbances (see Muller & Fayant, 2010, for a review), which can result in decreased academic performance on tasks that occur subsequent to upward comparison (Lyubomirsky et al., 2003).

Given the tendency for students to make upwards social comparisons, it is likely that using leaderboards in academic environments will prompt these upward comparisons. Additionally, research on leaderboards in the workplace has found that leaderboards encourage social comparisons among employees, so it is not unexpected that they would do so among students as well (Costa et al., 2013). Such comparisons could then result in poorer subsequent academic performance and a decrease in academic identification.

4. The current study

The current study investigates the impact of leaderboards on academic performance within the context of women and math skills. This particular context was selected because the effect of stereotype threat on women's math performance has been well-documented (see Nguyen & Ryan, 2008) and the ease with which sex can be indicated using only first names. Indeed, even very young children (3–3½ year olds) can reliably associate male and female names with images of boys and girls (Bauer & Coyne, 1997).

Based on the previous overview of the literature, the current study aims to explore whether or not viewing and interacting with leaderboards affects academic performance. Specifically, we compared the impact of male-dominated (i.e., a majority of male names, with males in higher positions) and female-dominated (i.e., a majority of female names, with females in higher positions) leaderboards on women's consequent performance on a math quiz and their academic self-concept, both in general and for math specifically.

Leaderboards are most commonly found within digital environments, such as online distance learning programs, edugames, and virtual classrooms. These digital environments have unique affordances (Bailenson et al., 2008). Unlike a physical classroom, a virtual environment classroom can be tailored to maximize learning outcomes for each individual student (Bailenson et al., 2008; McCall, Bailenson, Blascovich, & Beall, 2009). Unlike video conferencing, distractions from other students can be minimized (Bailenson et al., 2008). Compared to other forms of online learning, virtual classrooms provide greater immersion and richer, dynamic communication. Thus, it is unsurprising that the use of virtual classrooms is on the rise (see Cheryan, Meltzoff, & Kim, 2011; Duncan, Miller, & Jiang, 2012; Pellas, 2012, 2014; Wang & Burton, 2012), making virtual learning environments increasingly common and ushering in a need for research conducted within these contexts. While the current study does not take place within a true virtual classroom context (i.e., an environment in which an actual course is taught), it does use a virtual representation of a classroom environment that contains a number of cues meant to make the academic context of the study salient (see Figs. 3 and 4). This choice of setting allows very strict experimental control while still simulating an academic context. As it is possible that the relatively novel experience of navigating a virtual world may cause additional unwanted variance a third condition—no leaderboard—will be used to provide a control condition.

Previous findings indicate that there are two psychological processes, stereotype threat and social comparison, that might occur when women are exposed to leaderboards that draw attention to better or worse performance by members of their own sex. Each process would be expected to influence women's math performance, academic identification, and math identification in fairly different ways.

Stereotype threat research suggests that even minor cues can cause participants to experience stereotype threat and perform more poorly on stereotype-relevant tasks as a result (Inzlicht & Ben-Zeev, 2000). The stereotype applicable to this study is that "women are worse than men at math." Thus we would expect that the male-dominated leaderboard would cue the stereotype, whereas the female-dominated leaderboard would not.

H1a: Women exposed to the male-dominated leaderboard will perform more poorly on the subsequent math test than women exposed to the female-dominated leaderboard.

As the relevant stereotype only concerns a woman's mathematical performance, and disidentification only applies to the stereotyped domain (Cundiff et al., 2013), we would expect to see a drop in math identification, but not in overall academic identification.

Fig. 2. The female avatar used in the virtual classroom.

H2a: After completing the math test, women exposed to the male-dominated leaderboard will be lower in math identification than those exposed to the female-dominated leaderboard.

Social comparison theory, however, predicts a slightly different pattern of results. Being exposed to a gendered task, such as taking a math test, tends to make sex more salient, which then results in people processing and organizing new information along this dimension (Bem, 1981). Given that sex is already an important dimension for selecting targets for social comparison across a variety of domains (Dijkstra et al., 2008; Major & Forcey, 1985; Suls, Gaes, & Gastorf, 1979; Zanna, Goethals, & Hill, 1975), the heightened salience of sex makes it possible that the women in this experiment will compare themselves only with other women. Because upward social comparisons can result in decreased academic performance (Dijkstra et al., 2008; Lyubomirsky et al., 2003), we would expect to see women exposed to the female-dominated leaderboard perform more poorly than those exposed to the male-dominated leaderboard due to the impact of detrimental upward social comparisons with high-achieving women.

H1b: Women exposed to the female-dominated leaderboard will perform more poorly on subsequent math test than women exposed to the male-dominated leaderboard.

Given that upwards social comparison can influence domain identification (Dijkstra et al., 2008), one might expect to see decreases in both math identification and academic identification. However, research suggests that when an individual does not measure up to a standard in one area, they may choose to engage in defensive self-regulation (Muller & Fayant, 2010). One method of defensive regulation is to shift focus from the threatened domain to a higher-order domain in which the individual succeeds (Muller & Fayant, 2010). Thus, if women are engaging in upwards social comparison with the math leaderboard, we might expect to see a pattern such that identification with the specific domain in question (i.e., math) is decreased, but identification with a higher-order domain (i.e., academics in general) is improved.

H2b: After completing the math test, women exposed to the female-dominated leaderboard will be lower in math identification than those exposed to the male-dominated leaderboard or those in the control condition.

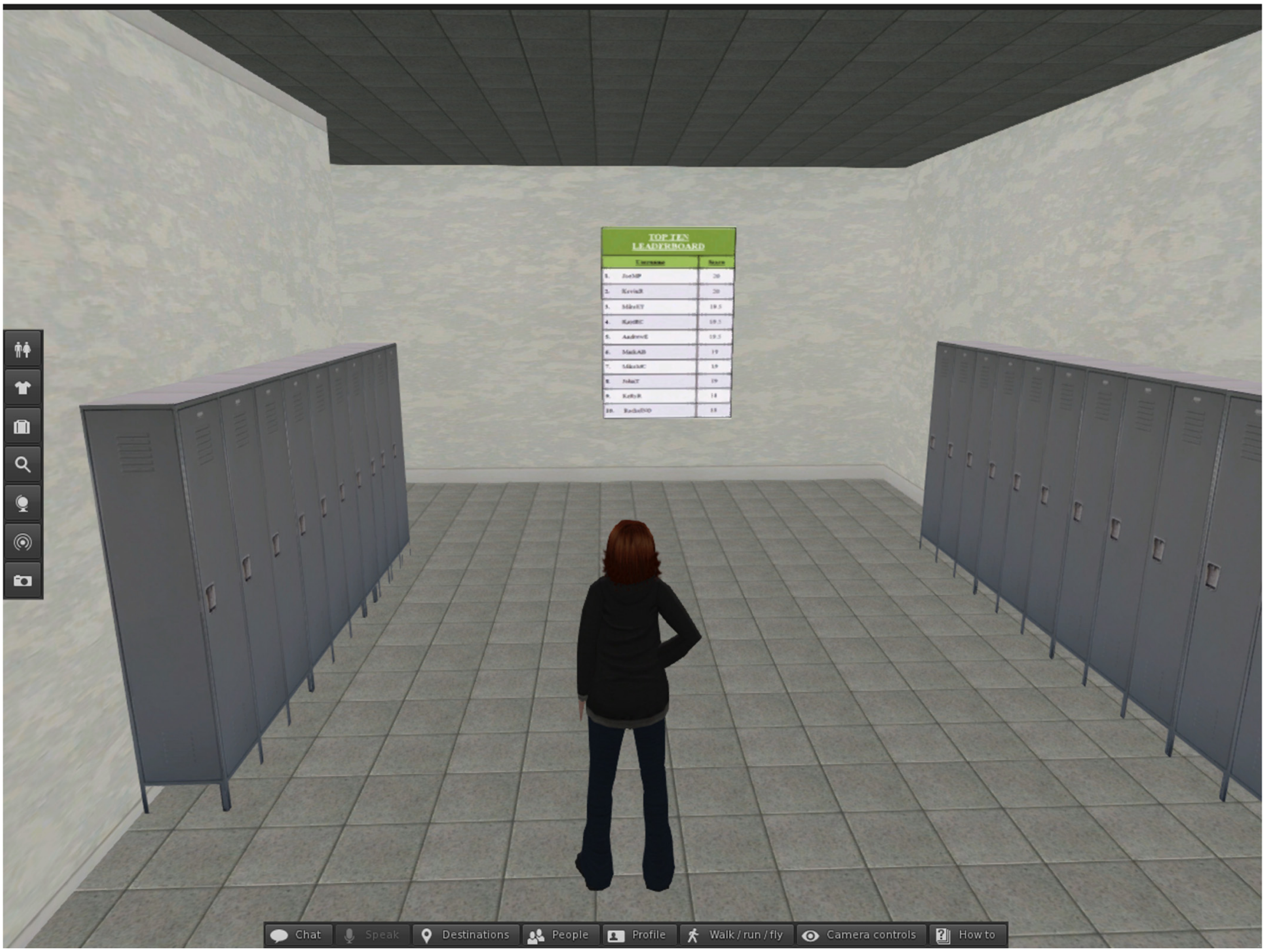

Fig. 3. The female avatar entering the room with the leaderboard.

H3: After completing the math test, women exposed to the female-dominated leaderboard will be higher in academic identification than those exposed to the male-dominated leaderboard or those in the control condition.

5. Method

5.1. Participants

Eighty women aged 18 to 41 ($M = 20.06$, $SD = 2.76$) were recruited from a student participant pool created from Communication courses at a large Midwestern university and received credit to participate in the study. Participants reported their race/ethnicity as: 83.8% Caucasian/European-American/White ($n = 67$); 7.5% Black/African/African-American ($n = 6$); 5% Asian/Asian-American ($n = 4$); 2.5% other ($n = 2$); and one participant did not respond.

5.2. Procedure

At least one week before the lab experiment, participants completed a pre-test consisting of demographic questions; measures assessing math identification and general academic identification; and several questions asking about their current and prior academic accomplishments, including the highest level of math completed. Upon arriving in the lab, each participant was seated in a single-occupancy room, where they could not see or interact with any other participants. After completing the consent process, they were informed that they would be participating in a study investigating the effectiveness of a virtual distance learning environment. This was done in order to mask the true purpose of the study from the participants. Participants were also told that they would be asked to evaluate their experiences of the virtual classroom representation and complete a brief math quiz within the virtual environment. Participants were randomly assigned to one of three conditions: the male-dominated leaderboard group (i.e., most and highest positions occupied by men), the female-dominated leaderboard group (i.e., most and highest positions occupied by women), and a control group that did not view a leaderboard. Fig. 1 shows a sample leaderboard.

Participants entered the virtual environment and took control of a female avatar, which remained identical across all conditions (see Fig. 2). Participants first completed a brief tutorial to help them learn how to navigate the virtual world and then interacted with a practice quiz consisting of opinion questions (e.g., what is your favorite color?) to ensure that they were familiar with how the quiz program looked and operated. This was done to ensure that performance on the math quiz was not influenced by a lack of familiarity with or understanding of the interface.

After completion of the tutorial, participants were asked to enter the experimental space, which had been set up to resemble a school hallway with lockers and a leaderboard at the end of the hall (see Fig. 3). Participants assigned to the male-dominated leaderboard condition saw a leaderboard where male names outnumbered female names and males held the majority of top places, whereas participants assigned to the female-dominated leaderboard condition saw a leaderboard where female names outnumbered male names and females held the majority of top places.

Participants were asked to register their names with the virtual leaderboard so that their quiz scores could be recorded; participants in the no leaderboard condition were merely told that their quiz scores would be recorded. Clicking on the leaderboard prompted participants to provide their names while simultaneously displaying the leaderboard clearly on the screen for the duration of the interaction, which lasted at least 10 s.

Participants were then instructed to examine the virtual classroom representation (see Fig. 4). Again, this was done in order to help conceal the true purpose of the study from the participants. Once they were familiar with the environment, they were instructed to sit in a test desk in the virtual environment (see Fig. 5). Clicking on a notebook on the desk started the math quiz. All participants had 15 min to complete the quiz and were permitted to use scratch paper, but not a calculator, to solve the problems.

Once participants had finished the in-world quiz, they were asked to complete a post-test focusing on their experiences within the virtual environment. This post-test included both distracter questions (e.g., How frequently did you think about something else while you were in the virtual classroom) and measures for the variables of interest. Participants did not receive any feedback regarding their scores on the test. After participants completed the post-test, they were debriefed, thanked for their time, and dismissed.

Fig. 4. The virtual classroom.

Fig. 5. The desk where the participant seated the avatar and took the math quiz.

5.3. Measures

5.3.1. Quiz score

Participants' math skills were assessed via an in-world quiz consisting of 20 multiple-choice questions. Questions on the quiz were taken from Graduate Record Examination (GRE) study materials provided by the Educational Testing Service. In an effort to minimize the impact of guessing, participants were told that skipping a question would not negatively impact their final score, but that getting a question wrong would result in the subtraction of ½ a point from their final score. For sample questions, please see Appendix A.

5.3.2. Math identification

Participants completed a self-assessment of their math identification on both the pre-test and immediately after completing the math quiz. Marsh's (1990) Academic Skill Self-Assessment Questionnaire (ASDQII) was used to measure math identification. The ASDQII asks participants to assess their performance in given academic areas, as well as to indicate how important doing well in the area is to them and how well they feel they do in the area compared to others. The participants completed items targeted toward math classes, resulting in an 8 item index measured on a 5-point Likert scale ranging from "Strongly Disagree" to "Strongly Agree" (Time 1 Cronbach's $\alpha = .92$; Time 2 $\alpha = .93$). Sample items include "I get good grades in math classes" and "I learn things quickly in math classes."

5.3.3. Academic identification

Participants completed a self-assessment of their academic identification on both the pre-test and immediately after completing the math quiz. Academic identification refers to how important academics *in general* are to the participant's identity (Osborne, 1997). The Academic Identification Questionnaire (Osborne, 1997) consists of 13 items measured on a 5-point scale ranging from "Strongly Disagree" to "Strongly Agree" (Time 1 Cronbach's $\alpha = .76$; Time 2 $\alpha = .82$). Sample items include "Being a good student is an important part of who I am" and "I often do my best work in school."

5.3.4. Math level

To control for pre-existing math skill, participants were asked in the pre-test to report the highest level of math class that they had successfully completed (algebra, geometry, pre-calculus, calculus).

6. Results

Means and standard deviations on the measures of interest for each group can be found in Table 1. Math identification at time 1, academic identification at time 1, and math level were all examined as possible covariates. Only significant covariates were retained for the analysis.

6.1. Performance on the math test

An ANCOVA was used to test for differences in scores across the three leaderboard conditions (male-dominated, female-dominated, and no leaderboard), with quiz score entered as the dependent variable and math level entered as a covariate. Math level was a significant covariate, $F(2, 76) = 8.04$, $p = .006$, $\eta^2_{\text{partial}} = .10$. Results showed that leaderboard condition had a significant impact on quiz score, $F(2, 76) = 3.45$, $p = .037$, $\eta^2_{\text{partial}} = .08$.

Planned pairwise contrasts revealed that participants in the female-dominated leaderboard condition ($M = 10.77$, $SD = 2.62$) performed significantly worse than participants in the male-dominated condition ($M = 12.46$, $SD = 2.12$, $t(76) = -2.62$, $p = .01$, $r = .29$). No other contrasts were significant. Thus, the stereotype threat hypothesis (H1a) was rejected and the social comparison hypothesis (H1b) was supported.

6.2. Math identification

An ANCOVA was used to test for differences between leaderboard conditions on post-test math identification. Math level and pre-test math identification were entered as covariates. Additionally, because the measure of identification was taken following the quiz, quiz score was also controlled for. Both pre-test math identification, $F(1, 74) = 173.48$, $p = .000$, $\eta^2_{\text{partial}} = .70$, and quiz score, $F(1, 74) = 4.54$, $p = .036$, $\eta^2_{\text{partial}} = .06$, were significant. Results show that leaderboard condition had no significant effect on post-test math identification, $F(2, 74) = .06$, $p = .947$, $\eta^2_{\text{partial}} = .00$. Thus, both hypotheses (H2a & H2b) were rejected.

6.3. Academic identification

An ANCOVA was used to test for differences between leaderboard conditions on post-test academic identification. Math level, quiz score, and pre-test academic identification score were entered as covariates. The only significant covariate was pre-test academic identification score, $F(1, 74) = 108.67$, $p = .000$, $\eta^2_{\text{partial}} = .60$. Leaderboard condition had a significant impact on post-test academic identification scores, $F(2, 74) = 5.53$, $p = .006$, $\eta^2_{\text{partial}} = .13$.

Planned pairwise contrasts revealed that participants in the female-dominated leaderboard condition ($M = 3.78$, $SD = .38$) had significantly higher post-test academic identification scores than those in either the male-dominated leaderboard condition ($M = 3.46$, $SD = .48$, $t(76) = -3.24$, $p = .004$, $r = .32$) or the control condition ($M = 3.40$, $SD = .54$, $t(76) = -2.72$, $p = .008$, $r = .30$). Thus, the social comparison hypothesis (H3) was supported.

7. Conclusions and discussion

In this study, female participants entered a virtual representation of a classroom and were exposed to either a male-dominated leaderboard, a female-dominated leaderboard, or no leaderboard and then asked to complete a math quiz. The results of the study found that women in the female-dominated leaderboard condition performed worse on the math quiz than those in the male-dominated leaderboard condition. Despite their objectively poorer performance, women in the female-dominated leaderboard condition demonstrated stronger academic identification than those in the control and male-dominated leaderboard conditions. There was no effect of leaderboard condition on math identification.

The results of this study suggest that the use of leaderboards in academic environments can, in some circumstances, affect academic performance. In this particular study, leaderboards appear to have inspired social comparison processes more so than the experience of stereotype threat. The female participants in this study appear to have made upward social comparisons with the high-achieving females in the female-dominated leaderboard condition, whereas participants in the male-dominated condition either did not make comparisons or made only downwards (i.e., self-enhancing) comparisons with the lower-achieving females on the board in that condition. Given that our sample contained only women, it is not possible to generalize these findings to men. Further research should continue to study these processes in the context of male students and minority students.

Social comparison processes also explain why general academic identification only increased for participants in the female-dominated leaderboard condition. When an individual feels they do not measure up to a standard in one area, as people often do after social comparisons along a specific dimension, they may choose to engage in defensive self-regulation (Muller & Fayant, 2010). One method of defensive regulation is to shift focus from the threatened domain to a higher-order domain in which the individual succeeds (Muller &

Table 1
Participant characteristics by condition.

	Male-dominated leaderboard	Female-dominated leaderboard	Control
n	26	31	23
Mean age (*SD*)	19.42 (1.36)	19.9 (1.30)	21 (4.64)
Average math level	Pre-calculus	Pre-calculus	Pre-calculus
Mean academic identification: pre-test (*SD*)	3.47 (.56)	3.57 (.43)	3.37 (.40)
Mean math identification: pre-test (*SD*)	3.1 (.78)	3.2 (.83)	3.33 (.82)

Note. There were no significant differences between groups for math level, pre-test academic identification, or pre-test math identification.

Fayant, 2010). In this case, the women who experienced a threat to their mathematical ability as a result of upward social comparison may have shifted focus from their math performance to their overall academic accomplishments (i.e., "It's okay that I'm not good at math because I am good at academics overall.")

This study found no result of leaderboard condition on math identification, however. This lack of influence on math identification may be the result of the timing of the measure, as it was taken after the quiz. Because of the salience of the quiz, participants' perception of their performance may have washed out any effects specific to the leaderboard treatment in relation to math identification.

7.1. Practical applications

Given that this is a single experimental study conducted within a highly-controlled environment, it is difficult to make any definitive statements regarding the common impact of using leaderboards within virtual learning environments, particularly given the mixed findings. Although female-dominated leaderboards may promote worse performance in women, they may also promote greater academic identification. This study does, however, provide an important lesson: it empirically demonstrates that the inclusion of leaderboards within educational environments can produce unexpected effects and that these effects can—under some circumstances—negatively impact student performance. Thus, until further research can be conducted, instructors would do well to be cautious in their use of leaderboards in virtual classrooms and other academic environments, especially when such leaderboards include specific, objective records of academic achievement such as quiz scores.

This study is also one of the few that have looked at such issues within a virtual academic context. The use of virtual classrooms is on the rise (see Duncan et al., 2012; Pellas, 2012, 2014; Wang & Burton, 2012), and they offer unique affordances not available in the traditional classroom (Bailenson et al., 2008). Future research should continue to investigate the potential of virtual classrooms and the various effects they may have on learning.

7.2. Future directions and limitations

Although we used an experimental design to maximize control and filter out the variables that often confound field studies of educational gamification, there are some limitations to this method. Participants were exposed to the leaderboard for a short period of time and experienced one learning assessment. Longitudinal controlled experiments are necessary to determine the ongoing effects of exposure to leaderboards as well as the interactive effects of one's own performance. Because we were maximizing control, participants in this study did not receive feedback on their performance. Future studies should examine how a student's positioning on the leaderboard—or absence from it—after assessment influences subsequent reactions to the leaderboard and academic performance.

Another limitation of this study is that the participants were not part of a real class and had not had any instruction within the virtual "classroom." This lack of a social element may have influenced our results, as the students had no real connection to any of the "people" named on the leaderboard. Future research should look at evaluating the impact of leaderboards within a more naturalistic virtual classroom setting.

Additionally, although this study's evidence strongly points to social comparison processes, participants in this study were not directly asked about their comparison processes to avoid sensitizing participants. Comparison processes are usually measured by either asking the participants to rate the target others as well as themselves (e.g., Mendes, Blascovich, Major, & Seery, 2001) or by asking the participants to rate themselves as compared to the target others (e.g., Sieverding & Koch, 2009). Future studies should attempt to directly assess social comparison behaviors in virtual classrooms. In the context of leaderboards and women's math performance, this would likely take the form of asking participants to either assess average female math ability after leaderboard exposure or to assess their own math ability as compared to the women they saw on the leaderboard. Another possible route for assessing social comparison effects would be to assess participants' pre- and post-test affect.

Finally, this study is somewhat limited due to its sample. The participants in this study are predominantly young, college-educated women. Although this allows this particular study to be quite applicable for understanding the use of virtual environments in higher education in the U.S., it is unclear whether or not the same findings would hold true for young girls (e.g., elementary aged), women in other cultures, and/or women with lower levels of formal education.

7.3. Conclusion

Despite these limitations, this study provides some of the first empirical evidence of the impact that leaderboards can have on students' academic performance. Although the use of new tools and techniques in the classroom certainly hold the potential to enhance students' educational experiences, they also, as demonstrated here, may have unanticipated negative influences. Given we are still in the process of understanding these effects, this study underscores the need for careful, empirical investigation of gamification and other trends in education before they are widely implemented in the classroom.

Appendix A. Math quiz example questions

Basic math

1. $15 - (6 - 4)(-2) = ?$
2. $(2 - 17)/5 = ?$

Algebra

1. $5x - 7 = 28, x = ?$
2. $10 - 5x = x + 30, x = ?$

Word problem

1. Two cars started from the same point and traveled in opposite directions for 2 h, at which time they were 208 miles apart. If one car traveled 8 miles per hour faster than the other car, what was the average speed of each car during the trip?

References

Aaronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29–46. http://dx.doi.org/10.1006/jesp.1998.1371.

Anderson, J. Q., & Rainie, L. (2012). Gamification: experts expect 'game layers' to expand in the future, with positive and negative results. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_Future_of_Internet_2012_Gamification.pdf.

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences, 17*, 102–141. http://dx.doi.org/10.1080/10508400701793141.

Bajdor, P., & Dragolea, L. (2011). The gamification as a tool to improve risk management in the enterprise. *Annales Universitatis Apulensis-Series Oeconomica, 13*, 574–583.

Bauer, P. J., & Coyne, M. J. (1997). When the name says it all: preschoolers' recognition and use of the gendered nature of common proper names. *Social Development, 6*, 271–291. http://dx.doi.org/10.1111/j.1467-9507.1997.tb00106.x.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: mechanisms, alleviation, and spillover. *Journal of Experimental Psychology, 136*, 256–276. http://dx.doi.org/10.1037/0096-3445.136.2.256.

Bem, S. L. (1981). Gender schema theory: a cognitive account of sex typing. *Psychological Review, 88*, 354–364. http://dx.doi.org/10.1037/0033-295X.88.4.354.

Boulet, G. (2012). Gamification: the latest buzzword and the next fad. *eLearn Magazine*. Retrieved from http://elearnmag.acm.org/archive.cfm?aid=2421596.

Butler, C. (2013). The effect of leaderboard ranking on players' perception of gaming fun. In A. A. Ozok, & P. Zaphiris (Eds.), *Online communities and social computing* (pp. 129–136). New York, NY: Springer.

Buunk, A. P., & Gibbons, F. X. (2007). Social comparison: the end of a theory and the emergence of a field. *Organizational Behavior & Human Decision Processes, 102*, 3–21. http://dx.doi.org/10.1016/j.obhdp.2006.09.007.

Camilleri, V., Busuttil, L., & Montebello, M. (2011). Social interactive learning in multiplayer games. In M. Ma, A. Oikonomou, & L. C. Jain (Eds.), *Serious games and edutainment applications* (pp. 481–501). London, England: Springer-Verlag.

Charsky, D. (2010). From edutainment to serious games: a change in the use of game characteristics. *Games & Culture, 5*, 177–198. http://dx.doi.org/10.1177/1555412009354727.

Cheryan, S., Meltzoff, A. N., & Kim, S. (2011). Classrooms matter: the design of virtual classrooms influences gender disparities in computer science classes. *Computers & Education, 57*, 1825–1835. http://dx.doi.org/10.1016/j.compedu.2011.02.004.

Chou, Y. (2013). A comprehensive list of 90+ gamification cases with ROI stats. *Yu-Kai Chou & Gamification*. Retrieved from http://www.yukaichou.com/gamification-examples/gamification-stats-figures/#.UzYtafldWVp.

Cohen, A. M. (2011). The gamification of education. *Futurist, 45*, 16–17.

College Board. (2012). *2012 College-bound seniors total group profile report*. Retrieved from http://media.collegeboard.com/digitalServices/pdf/research/TotalGroup-2012.pdf.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*, 661–686. http://dx.doi.org/10.1016/j.compedu.2005.09.001.

Costa, J. P., Wehbe, R. R., Robb, J., & Nacke, L. E. (2013). Time's up: studying leaderboards for engaging punctual behaviour. In *Gamification 2013: Proceedings of the first international conference on gameful design, research, and applications* (pp. 1–8). Retrieved from http://www.researchgate.net/publication/257519212_Time's_Up_Studying_Leaderboards_For_Engaging_Punctual_Behaviour/file/e0b4952558a11dc085.pdf.

Cundiff, J. L., Vescio, T. K., Loken, E., & Lo, L. (2013). Do gender-science stereotypes predict science identification and science career aspirations among undergraduate science majors? *Social Psychology of Education, 16*, 541–554. http://dx.doi.org/10.1007/s11218-013-9232-8.

de-Marcos, L., Domínguez, A., Saenz-de-Navarrete, J., & Pagés, C. (2014). An empirical study comparing gamification and social networking on e-learning. *Computers & Education, 75*, 82–91. http://dx.doi.org/10.1016/j.compedu.2014.01.012.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining "gamification". In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9–15). http://dx.doi.org/10.1145/2181037.2181040.

Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A. P., & van der Zee, Y. G. (2008). Social comparison in the classroom: a review. *Review of Educational Research, 78*, 828–879. http://dx.doi.org/10.3102/0034654308321210.

Domínguez, A., Saenz-de-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J. J. (2013). Gamifying learning experiences: practical implications and outcomes. *Computers & Education, 63*, 380–392. http://dx.doi.org/10.1016/j.compedu.2012.12.020.

Duncan, I., Miller, A., & Jiang, S. (2012). A taxonomy of virtual worlds usage in education. *British Journal of Educational Taxonomy, 43*, 949–964. http://dx.doi.org/10.1111/j.1467-8535.2011.01263.x.

Ellison, G., & Swanson, A. (2010). The gender gap in secondary school mathematics at high achievement levels: evidence from the American mathematics competitions. *Journals of Economic Perspectives, 24*, 109–128. http://dx.doi.org/10.1257/jep.24.2.109.

Farzan, R., MiMicco, J. M., Millen, D. R., Dugan, C., Geyer, W., & Brownholz, E. A. (2008). Results from deploying a participation incentive mechanism within the enterprise. In *CHI '08 proceedings of the SIGCHI conference on human factors in computing systems* (pp. 563–572). http://dx.doi.org/10.1145/1357054.1357145.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140. http://dx.doi.org/10.1177/001872675400700202.

Hamari, J. (2013). Transforming homo economicus into homo ludens: a field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic Commerce Research and Applications, 12*, 236–245. http://dx.doi.org/10.1016/j.elerap.2013.01.004.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? A literature review of empirical studies on gamification. In *Proceedings of the 47th international conference on system sciences*. Retrieved from http://www.researchgate.net/publication/256743509_Does_Gamification_Work__A_Literature_Review_of_Empirical_Studies_on_Gamification/file/9c960523b36be81843.pdf.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*, 41–45. http://dx.doi.org/10.1126/science.7604277.

Hoorens, V., & Van Damme, C. (2012). What do people infer from social comparisons? Bridges between social comparison and person perception. *Social and Personality Psychology Compass, 6*, 607–618. http://dx.doi.org/10.1111/j.1751-9004.2012.00451.x.

Huguet, P., & Régner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology, 99*, 545–560. http://dx.doi.org/10.1037/0022-0663.99.3.545.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11*, 365–371. http://dx.doi.org/10.1111/1467-9280.00272.

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: examining the influence of emotion regulation. *Journal of Experimental Social Psychology, 137*, 691–705. http://dx.doi.org/10.1037/a0013834.

Jones, J. B. (2010). Gamifying homework. *The Chronicle of Higher Education: ProfHacker*. Retrieved from http://chronicle.com/blogs/profhacker/gamifying-homewor/28407.

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. San Francisco, CA: John Wiley & Sons Inc.

Kaya, T. (2010). A "stealth assessment" turns to video games to measure thinking skills. *The Chronicle of Higher Education*. Retrieved from http://chronicle.com/article/A-Stealth-Assessment-Turns/125276/.

Keller, J. (2002). Blatant stereotype threat and women's math performance: self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles, 47*, 193–198. http://dx.doi.org/10.1023/A:1021003307511.

Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior, 35*, 179–188.

Landers, R. N., & Callan, R. C. (2011). Casual social games as serious games: the psychology of gamification in undergraduate education and employee training. In M. Ma, A. Oikonomou, & L. C. Jain (Eds.), *Serious games and edutainment applications* (pp. 399–423). London, England: Springer-Verlag.

Lee, J. J., & Hammer, J. (2011). Gamification in education: what, how, why bother? *Academic Exchange Quarterly, 15*, 146–150.

Lyubomirsky, S., Kasri, F., & Zehm, K. (2003). Dysphoric rumination impairs concentration on academic tasks. *Cognitive Therapy and Research, 27*, 309–330. http://dx.doi.org/10.1023/A:1023918517378.

Major, B., & Forcey, B. (1985). Social comparisons and pay evaluations: preferences for same-sex and same-job wage comparisons. *Journal of Experimental Social Psychology, 21*, 393–405. http://dx.doi.org/10.1016/0022-1031(85)90038-1.
Marsh, H. W. (1990). The structure of academic self-concept: the Marsh/Shavelson model. *Journal of Educational Psychology, 82*, 623–636. http://dx.doi.org/10.1037/0022-0663.82.4.623.
Marx, D. M., & Roman, J. S. (2002). Female role models: protecting women's math test performance. *Personality and Social Psychology Bulletin, 28*, 1183–1193. http://dx.doi.org/10.1177/01461672022812004.
McCall, C., Bailenson, J. N., Blascovich, J., & Beall, A. C. (2009). Leveraging collaborative virtual environment technology for inter-population research on persuasion in a classroom setting. *Presence: Teleoperators & Virtual Environments, 18*, 361–369.
McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology, 39*, 83–90. http://dx.doi.org/10.1016/S0022-1031(02)00513-9.
Mendes, W. B., Blascovich, J., Major, B., & Seery, M. (2001). Challenge and threat responses during downward and upward social comparison processes. *European Journal of Social Psychology, 31*, 477–497. http://dx.doi.org/10.1002/ejsp.80.
Muller, D., & Fayant, M.-P. (2010). On being exposed to superior others: consequences of self-threatening upward social comparisons. *Social and Personality Psychology Compass, 4*, 621–634. http://dx.doi.org/10.1111/j.1751-9004.2010.00279.x.
Muntean, C. I. (2011). Raising engagement in e-learning through gamification. In *Proceedings of the 6th international conference on virtual learning* (pp. 323–329). Retrieved from http://icvl.eu/2011/disc/icvl/documente/pdf/met/ICVL_ModelsAndMethodologies_paper42.pdf.
Nguyen, H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*, 1314–1334. http://dx.doi.org/10.1037/a0012702.
Osborne, J. W. (1997). Identification with academics and academic success among community college students. *Community College Review, 25*, 59–67. http://dx.doi.org/10.1177/009155219702500105.
Pellas, N. (2012). Towards a beneficial formalization of cyber entities' interactions during the e-learning process in the virtual world of "Second Life." In P. Renna (Ed.), *Production and manufacturing system management: Coordination approaches and multi-site planning* (pp. 242–277) Hershey, PA: Engineering Science Reference.
Pellas, N. (2014). The influence of computer self-efficacy, metacognitive self-regulation and self-esteem on student engagement in online learning programs: evidence from the virtual world of Second Life. *Computers in Human Behavior, 35*, 157–170. http://dx.doi.org/10.1016/j.chb.2014.02.048.
Pronin, E., Steele, M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: women and mathematics. *Journal of Experimental Social Psychology, 40*, 152–168. http://dx.doi.org/10.1016/S0022-1031(03)00088-X.
Reeves, B., & Read, J. L. (2009). *Total engagement*. Boston, MA: Harvard Business Press.
Salter, A. (2011a). Games in the classroom (part 1). *The Chronicle of Higher Education: ProfHacker*. Retrieved from http://chronicle.com/blogs/profhacker/games-in-the-classroom-part-1/35596.
Salter, A. (2011b). How to 'gamify' your class website. *The Chronicle of Higher Education: ProfHacker*. Retrieved from http://chronicle.com/blogs/profhacker/how-to-gamify-your-class-website/31332.
Salter, A. (2011c). Lessons learned in playful game design. *The Chronicle of Higher Education: ProfHacker*. Retrieved from http://chronicle.com/blogs/profhacker/lessons-learned-in-playful-game-design/33876.
Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*, 440–452. http://dx.doi.org/10.1037/0022-3514.85.3.440.
Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: the role of stereotype endorsement in women's experience in the math domain. *Sex Roles, 50*, 835–850. http://dx.doi.org/10.1023/B: SERS.0000029101.74557.a0.
Schmader, T., Major, B., & Gramzow, R. H. (2001). Coping with ethnic stereotypes in the academic domain: perceived injustice and psychological disengagement. *Journal of Social Issues, 57*, 93–111. http://dx.doi.org/10.1111/0022-4537.00203.
Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: their effects on women's performance. *Journal of Experimental Social Psychology, 39*, 68–74. http://dx.doi.org/10.1016/S0022-1031(02)00508-5.
Sieverding, M., & Koch, S. C. (2009). (Self-) evaluation of computer competence: how gender matters. *Computers & Education, 52*, 696–701. http://dx.doi.org/10.1016/j.compedu.2008.11.016.
Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–28. http://dx.doi.org/10.1006/jesp.1998.1373.
Steele, C. M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613. http://dx.doi.org/10.1037/0003-066X.52.6.613.
Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: the psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology, 34*, 379–440. http://dx.doi.org/10.1016/S0065-2601(02)80009-0.
Suls, J., Gaes, G., & Gastorf, J. (1979). Evaluating a sex-related ability: comparison with same-, opposite-, and combined-sex norms. *Journal of Research in Personality, 13*, 294–304. http://dx.doi.org/10.1016/0092-6566(79)90020-5.
Taylor, C. A., Lord, C. G., McIntyre, R. B., & Paulson, R. M. (2011). The Hillary Clinton effect: when the same role model inspires or fails to inspire improved performance under stereotype threat. *Group Processes & Intergroup Relations, 14*, 447–459. http://dx.doi.org/10.1177/1368430210382680.
Thompson, M., & Sekaquaptewa, D. (2002). When being different is detrimental: solo status and the performance of women and racial minorities. *Analyses of Social Issues and Public Policy, 2*, 183–203. http://dx.doi.org/10.1111/j.1530-2415.2002.00037.x.
Verkuyten, M., & Thijs, J. (2004). Psychological disidentification with the academic domain among ethnic minority adolescents in the Netherlands. *British Journal of Educational Psychology, 74*, 109–125. http://dx.doi.org/10.1348/000709904322848842.
Wang, F., & Burton, J. (2012). Second Life in education: a review of publications from its launch to 2011. *British Journal of Educational Technology, 44*, 357–371. http://dx.doi.org/10.1111/j.1467-8535.2012.01334.x.
Wehrens, M. J., Buunk, A. P., Lubbers, M. J., Dijkstra, P., Kuyper, H., & van der Werf, G. P. (2010). The relationship between affective response to social comparison and academic performance in high school. *Contemporary Educational Psychology, 35*, 203–214. http://dx.doi.org/10.1016/j.cedpsych.2010.01.001.
Woodcock, A., Hernandez, P. R., Estrada, M., & Schultz, P. W. (2012). The consequences of chronic stereotype threat: domain disidentification and abandonment. *Journal of Personality and Social Psychology, 103*, 635–646. http://dx.doi.org/10.1037/a0029120.
Xie, Y., & Shauman, K. A. (2003). *Women in science: Career processes and outcomes*. Cambridge, MA: Harvard University Press.
Zanna, M. P., Goethals, G. R., & Hill, J. F. (1975). Evaluating a sex-related ability: social comparison with similar others and standard setters. *Journal of Experimental Social Psychology, 11*, 86–93. http://dx.doi.org/10.1016/S0022-1031(75)80013-8.