



OPEN ACCESS

EDITED BY

Ana Mercedes Vernia-Carrasco,
University of Jaume I, Spain

REVIEWED BY

John R. Finnegan,
University of Minnesota Twin Cities,
United States
Andrea Vial,
New York University Abu Dhabi,
United Arab Emirates

*CORRESPONDENCE

Olga Khokhlova
✉ o.khokhlova@mdx.ac.ae

†These authors have contributed equally to this work and share first authorship

RECEIVED 03 February 2023

ACCEPTED 11 July 2023

PUBLISHED 27 July 2023

CITATION

Khokhlova O, Lamba N and Kishore S (2023) Evaluating student evaluations: evidence of gender bias against women in higher education based on perceived learning and instructor personality. *Front. Educ.* 8:1158132. doi: 10.3389/educ.2023.1158132

COPYRIGHT

© 2023 Khokhlova, Lamba and Kishore. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Evaluating student evaluations: evidence of gender bias against women in higher education based on perceived learning and instructor personality

Olga Khokhlova^{1*†}, Nishtha Lamba^{1†} and Sameer Kishore²

¹Social Psychology Research Lab, Middlesex University Dubai, Dubai, United Arab Emirates, ²VR^X Lab, Middlesex University Dubai, Dubai, United Arab Emirates

Given student evaluations are an integral part of academic employment and progression in higher education, it is crucial to explore various biases amongst students that may influence their ratings. Several studies report a clear gender bias in student evaluation where male instructors receive significantly higher ratings as compared to female instructors. However, there is very limited research about gender biases in underrepresented samples such as South Asia and the Middle East. We examined whether perception of male and female instructors differed in terms of how they facilitate learning and level of engagement, using an experimental design. Six hundred and seventy-one university students were asked to watch a video of an online lecture on psychology, delivered by either a male or female lecturer, after which they were asked to evaluate their experience and instructor personality characteristics. To ensure consistency across content, tone, delivery, environment, and overall appearance, photorealistic 3D avatars were used to deliver the lectures. Only gender as a factor was manipulated. Given the racial representation in the region, a total of four videos were developed representing males ($n = 317$) and females ($n = 354$) of White and South Asian race. Overall, male instructors scored significantly higher in variables representing personality characteristics such as enthusiasm and expressiveness compared to female instructors. Participants did not however view male and female instructors to be different in terms of presentation and subject knowledge. Findings related to facilitating learning suggest that male instructors were perceived to have made instructions more interesting, kept participants' attention for longer, and were more interesting compared to female instructors. In terms of engagement, male instructors were perceived to be more expressive, enthusiastic, and entertaining, compared to female instructors. Given the experimental design, these findings can clearly be attributed to gender bias, which is also in line with previous research. With an underrepresented sample, an online platform delivery, and inclusion of multiple races, these findings significantly add value to the current literature regarding gender stereotypes in higher education. The results are even more concerning as they provide strong evidence of gender bias which may contribute to subconscious discrimination against women academics in the region.

KEYWORDS

underrepresented samples, teaching perception, gender bias, discrimination, online teaching

1. Introduction

1.1. Gender bias in higher education

Gender bias in higher education has been widely reported and studied, with a particular emphasis on the experience of women in academia, but still remains a persistent issue. Despite various interventions aimed at mitigating gender inequality in higher education, such as the Athena Swan charter in the United Kingdom (Ovseiko et al., 2017), their overall impact has been less than satisfactory, particularly in the STEM disciplines (O'Connor, 2020). While certain initiatives, such as setting academic gender quotas, have had some degree of success in terms of hiring (Park, 2020), it is largely evident that women in academia continue to face significant barriers (Mama, 2003; Klein, 2016). In addition to the discrimination faced by women in the hiring process, selection for promotion, or assignment of leadership roles, several studies have also shown a bias in the evaluation of their teaching and professional contributions by Feldman (1993), Hamermesh and Parker (2005), Fan et al. (2019), and Renström et al. (2021). The issue of gender bias against women instructors is particularly relevant in higher education, where student evaluations of teachers (SET) are considered a significant factor in career advancement and promotion decisions. Usage of SETs as a metric for evaluating faculty performance has been repeatedly shown to be problematic in the literature (Wachtel, 1998; Zabaleta, 2007; Spooren et al., 2013; Boring et al., 2016) mainly addressing how the results of these evaluations are significantly biased against female instructors and somehow even influence how students rate objective aspects of teaching, such as the time taken to grade assignments (Kreitzer and Sweet-Cushman, 2021). While some researchers have attempted to provide alternatives (Hornstein, 2017), SETs still continue to be used by academic institutions to assess faculty performance and student satisfaction with modules. Please note that throughout the paper and in referenced literature, the term “gender” is used as a binary label. We acknowledge the complexity of the term as well as the possibility that binarization may not be accurate. However, for the purposes of this study we have used the term in the same way as used in the literature, without any other meaning attached.

Numerous studies have shown evidence of gender bias in student evaluations of instructors, with male instructors consistently receiving higher ratings compared to female instructors (Basow, 1995; Centra and Gaubatz, 2000), despite evidence of equal learning outcomes (Boring, 2017). In higher education, male instructors are often assumed to possess greater expertise in their subject matter, while female instructors are expected to prove themselves, an effect clearly illustrated when male instructors are often referred to as “professors” while female instructors are called “teachers” (Miller and Chamberlin, 2000; Renström et al., 2021). Furthermore, the interaction between instructors and students has an impact on evaluations, with significant differences noted between male and female instructors (Sinclair and Kunda, 2000). Negative feedback from female instructors was viewed as a sign of less competence, while the impact of negative feedback from male instructors was not as pronounced. Physical attributes completely unrelated to actual academic and teaching competence have been shown to have a significant impact on evaluations. Although some studies have

found no differences in evaluations based on gender (Tatro, 1995), other studies reveal that factors such as perceived beauty can impact both male and female lecturers (Hamermesh and Parker, 2005). Bennett (1982) showed that female lecturers were rated higher if they exhibited stereotypical feminine behaviors, while Babin et al. (2020) found that women faculty were rated more favorably if they were perceived to be more attractive, but this effect was diminished in an online setting.

In contrast, Tatro (1995) showed that female instructors received higher ratings in SETs than male instructors. In fact, a recent meta analysis which summarized findings from past 20 years of data showed that women scientists in maths-intensive fields have not largely experienced discrimination (Ceci and Williams, 2011). Additionally, it was found based on a hypothetical hiring scenario that female academics received preference in biology, engineering, economics, and psychology fields, suggesting that perhaps there is a shift in varied academic settings (Williams and Ceci, 2015). However, it is important to note that these studies have been administered in developed nations in the West.

1.2. Theoretical framework—shifting standards model

Stereotyped gender biases often emerge from having different reference points and evaluative standards for males and females in society, both in professional and personal spaces (Biernat and Manis, 1994). This varied lens of expectation could affect how male and female lecturers are perceived in the classroom. For example, women are expected to be more communal and men are expected to be more agentic. In addition, incongruent descriptive and prescriptive gender stereotypes may lead to further negative ratings. For example, women are expected to be more nurturing and communal, so female lecturers who appear to be agentic and assertive could be again judged more harshly (Rudman and Glick, 2001). According to the shifting standards model, an individual's assessment of members from stereotyped groups are influenced by shifting or flexible standards. That is, the same performance is often evaluated differently depending on the gender of the individual, with lower expectations for stereotypically disadvantaged groups, such as women in male-dominated domains like academia (Fuegen et al., 2004).

In the current study, we explore the issue of gender bias against women instructors in higher education by observing students' perceptions of learning and engagement using avatars. The study uses 3D virtual characters, or avatars, to deliver a pre-recorded online lecture presented via video in order to examine if these well-documented biases are also carried over in an online setting, and the implications of these findings for the design and use of avatars in higher education.

We anticipate unconscious biases to affect participants' judgements leading to gender bias against female instructors. More specifically, we would predict that, despite identical content delivery, avatars representing women instructors might be evaluated more critically than those representing men. This bias could manifest in various ways; for example, the same content presented by a male and female avatar might be perceived as

more interesting when delivered by the male, reflecting the harsher standards applied to women due to societal stereotypes. Conversely, if a female instructor avatar were to perform exceptionally well, her competence might be downplayed or seen as an exception, rather than the norm for women instructors.

The theories stated above supported the conceptualization of the research on gender bias and facilitated the design of the avatars. To our best knowledge while keeping theoretical framework and previous literature in mind we created avatars who were not overly feminine or masculine, reducing the potential impact of stereotyped gender biases on the participants' evaluations. While we expect unconscious biases to still influence the participants, we ensured not priming the participants with avatars appearing agentic, being extremely feminine or masculine in clothing, or having overly feminine or masculine body language.

1.3. Simulated vs. observed SET design

Within a typical educational setting, evaluation of instructors usually hinges upon end-of-year reviews. These evaluations, while useful to some extent, might not be sufficient to provide comprehensive insights into the underlying biases against instructors, largely due to their retrospective nature and dependence on students' subjective opinions. Alternatively, the use of avatars offers a more effective method for bias analysis. Avatars can be customized to reflect various demographic characteristics, facilitating a controlled environment in which biases can be isolated and measured in detail. This method allows for real-time monitoring and assessment, rather than retrospective evaluations, and provides a more objective metric for measuring biases, circumventing the typical cognitive biases and memory distortions associated with end-of-year evaluations. By presenting the same content through different avatars, any disparities in student responses can be attributed to perceptions of the avatar's characteristics rather than the instructional material itself, which provides an understanding of the nuanced impact of instructor identity on student perceptions and outcomes.

Additionally, using avatars to study these types of biases instead of evaluating observational data allows us to sidestep the bias associated with grading. Studies have shown that student evaluation of lecturers might be influenced by the grades they receive, with higher grades often translating into more positive evaluations (Stroebe, 2020; Berezvai et al., 2021). This introduces a substantial bias, as it conflates academic performance and favorability of instructor, undermining the objective assessment of teaching effectiveness. By using avatars, the personal relationship and familiarity aspects between students and instructors are essentially nullified. Each interaction is independent, and thus devoid of any previous personal encounters or grade-induced biases. This allows for a more accurate evaluation of the effect of the instructor's perceived identity. Hence, the use of avatars can result in an effective form of bias detection that is less tainted by variables like past grading history or relationship with the student.

Although there are a fair number of studies which have examined post-course evaluation, they have not focused on gender differences and have often been non-experimental in nature. These

studies either compared the evaluation of the course with the student demographic variables (Marsh, 1980) or investigated the affective aspect of student assessment of teaching, identifying domains important to teaching effectiveness (see Dziuban and Moskal, 2011). Very few studies have looked at post-course evaluations to compare the ratings of male and female intruders. Those that have done so have tended to be primarily interested in how the stereotypes found in student responses relate to actual knowledge and exam results (Boring, 2017). However, it should be noted that the research was carried out in face-to-face classes. Therefore, even when there was a random allocation of students to different instructors, the degree of experimental control of variables such as lecture content, physical characteristics, and clothing of the lecturer, body language, and basic attention was insufficient. Our aim was to exclude as many confounding variables as possible, so it was conducted with online avatars who allowed extensive experimental controls on these variables.

Another alternative to using avatars for observing biases is to use videos of actual instructors (Gillies, 2008; Chisadza et al., 2019). While videos can indeed provide a degree of control, specifically in the content being delivered, they lack the customization that avatars provide. A given video of an instructor is bound to the specific characteristics and attributes of that particular individual at that particular time. This is especially critical when comparing a specific factor, such as gender, where there might be several other differences in videos of male and female instructors, apart from just their gender, such as age, clothing, body language and movement.

1.4. Studying biases in an online environment

This difference in evaluation of instructors based on an online setting as compared to the in-person setting is of interest, given how teaching delivery has seen a paradigm shift since the COVID-19 pandemic (Bao, 2020; Rapanta et al., 2020). Studies have previously used online simulated learning environments to study the existence of gender bias, however there is a lack of consensus given the multi-faceted nature of the issue as well as the complexity of representation in the online environment. In their study, MacNell et al. (2015) manipulated the gender of the instructor by changing their online name in all communication with students including on message boards and emails and found the male "identity" to be rated significantly higher, regardless of the instructor's actual identity. On the other hand, students gave higher ratings to female instructors when they were shown a video lecture (Chisadza et al., 2019) even though the delivered content was exactly the same. Further observing evaluation of online instructors, the usage of pedagogical agents, virtual avatars aimed to "facilitate learning in computer-mediated learning environments" (Baylor and Kim, 2005), has been investigated. Numerous factors, including ethnicity (Baylor, 2005; Kim and Wei, 2011), gender, realism of the avatar, and delivery style (Baylor and Kim, 2004), have been analyzed with a typical tendency for students to favor instructors of the same gender as themselves (review in Kim and Baylor, 2016). Although the emphasis when designing and evaluating pedagogical agents is typically on the development and understanding of

student preferences rather than examining potential biases in student evaluations. In addition, the type of content that the pedagogical agent is meant to deliver is quite distinct to that of an instructor. The role of pedagogical agents is typically seen as a supplementary tool for student learning rather than a replacement for the instructor.

The increasing use of technology in higher education presents new challenges and opportunities for addressing gender bias against women instructors. Virtual avatars, in particular, have the potential to support more inclusive and engaging learning environments as they allow for greater control over the teaching environment and provide a standardized method for delivering instruction, reducing the impact of extraneous factors such as physical appearance, but they also raise questions about the potential for gender biases to persist and even amplify in these digital environments. The issue of gender bias in online learning environments, particularly when instructors are represented by avatars in an online lecture, has received relatively little attention. There is a critical need to understand the impact of gender bias on women instructors in online learning environments and to explore the potential for online avatars to counteract these biases and promote equal educational opportunities for all instructors.

1.5. Sample characteristics

The demographic composition of the UAE is a significant factor to consider. Data from the International Migrant Stock, provided by the UN Department of Economic and Social Affairs (2019), reveals that 88% of the UAE's population consists of expatriates, with 59% originating from South Asian countries. Therefore, this study aims to add to the generalizability of gender bias research recruiting university students, the majority of whom will be of South Asian cultural background, which is largely suggestive of the country's demographic landscape. In order to have a more realistic demographic representation in the region we have included instructors belonging to White and South Asian backgrounds. Although we examine differences based on gender, in this way race was also manipulated in the generated videos, primarily in order to provide a balanced representation to the participants according to the regional demographics.

Given the race manipulation, it is important to consider imbalance in evaluations based on race as well. Reid (2010) conducted a study analysing anonymous teaching evaluations sourced from [RateMyProfessors.com](https://www.ratemyprofessors.com), to measure impact of race (including Black, White, Asian, and Latino) and gender on faculty evaluations. Findings indicated that faculty from racial minority groups, specifically Black and Asians, received less positive reviews than their White counterparts regarding overall quality, helpfulness, and clarity. Interaction between student and teacher characteristics as reported in various studies do not always exhibit consistency. For instance, Chisadza et al. (2019) found that black students evaluated black lecturers lower as compared to white lecturers. To our knowledge, ours is the first study that also includes South Asian instructors when studying gender bias. A greater representation of lecturers and participants from countries which are usually underrepresented in research is an important strength

of the study. With a huge expatriate population, we expect gender bias against female lecturers in higher education in the UAE to be similar to previous research.

Furthermore, rather than evaluating instructors on multiple factors, the two main considerations made here are related to facilitating learning and student engagement. Facilitating learning is a critical aspect of teaching and involves creating a positive and engaging learning environment that supports student's development and success, while engagement alludes to how effectively the instructor is able to hold the students' attention and interest during the lecture. It was hypothesized that participants will show a gender bias in evaluations, such that they will evaluate female virtual instructors more negatively than male virtual instructors. The experimental design presented here offers strong internal and ecological validity allowing for an efficient way to study these biases. The subsequent sections describe the design of the study, including a description of how the avatars were generated, followed by a discussion of the results, implications and directions of future work using this platform that has been developed.

2. Methodology

2.1. Participants

Priori analysis using G*Power suggested that to achieve a desired power of 0.80, effect size 0.3, and $p < 0.05$ for two tailed independent t -tests, we required a minimum of 176 participants per group. We followed this guideline to establish a minimum sample size.

Our sample consisted of 691 participants being recruited through convenience sampling, which was considered suitable due to easy accessibility and cost-effectiveness. We did not have incomplete responses as we had included forced responses. Twenty participants were excluded from the final sample as they were High School students (under 17), gave impossible responses, did not qualify due to duplicate cases, did not answer at least two of our four attention questions, or simply did not consent to the study. The overall sample comprised of 671 participants (Mage = 20.12, SDage = 2.38, age range = 17–43). Of these, 74.5% were female ($n = 500$), 23.70% were male ($n = 159$), and 1.8% ($n = 12$) preferred not to reveal their gender. The majority of participants reside in the United Arab Emirates (83.7%, $n = 561$) and are South Asians (63.5%, $n = 426$).

The participants had diverse levels of education, consisting of individuals with a Doctoral-Level Qualification (0.1%), Postgraduate qualification (2.8%), Undergraduate qualification (50.7%), and current Undergraduate students with a High School Diploma (46.30%). As per the inclusion criteria, all participants had been in a classroom setting within the last 5 years, and 85.20% were currently enrolled as students. 32.30% of participants were psychology majors ($n = 217$), 8.90% studied computer science ($n = 60$), 7.20% studied business administration ($n = 48$), 6.30% studied engineering ($n = 42$), 5.40% studied medicine ($n = 36$), and the rest (39.9%, $n = 268$) were majoring in courses such as architecture, finance, law, accounting, marketing, management, and advertising.

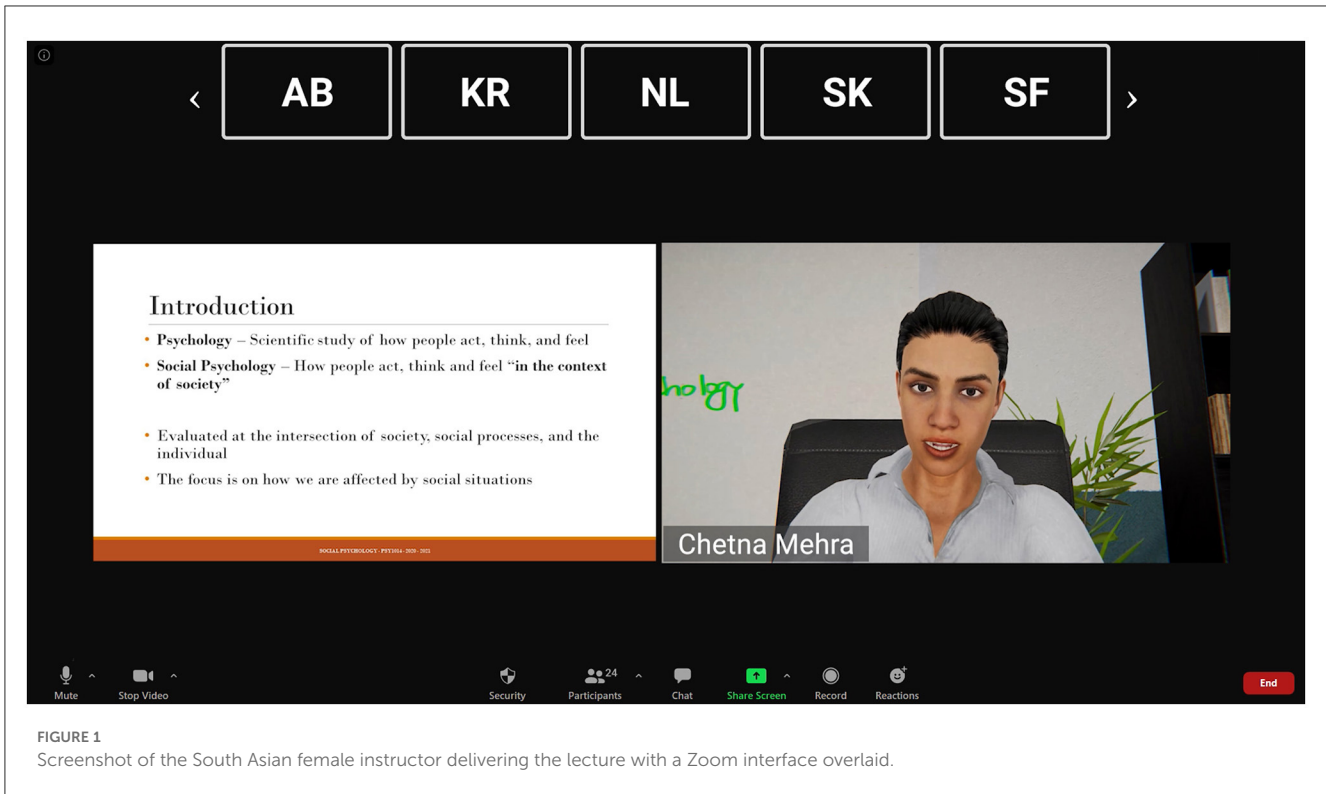


FIGURE 1
Screenshot of the South Asian female instructor delivering the lecture with a Zoom interface overlaid.

2.2. Procedure and design

To observe differences in student evaluation, the researchers created two videos where a virtual character (an animated avatar) delivered a 5 min fragment of a social psychology lecture. Introduction to social psychology was chosen for two reasons. Firstly, it requires no prior knowledge and is quite accessible to a larger audience. Secondly, research suggests that women are more discriminated against in academia when teaching STEM subjects (Babin and Hussey, 2023). Based on literature, we would expect the disparity in evaluations to be even larger in a STEM subject, we chose psychology in order to diminish that difference, thus making it a stronger test of the hypothesis. The research was not preregistered.

Given the ethnic representation in the region, South Asian and White male and female avatars were created using software Character Creator 3 by Reallusion.¹ All videos were identical in terms of the script those characters articulated, the background of the room and their clothes. The researchers were careful not to prime participants with information about these characters except for the manipulated factors (gender and race). Thus, participants watched one of four videos (South Asian Male, South Asian Female, White Male, and White Female), and for the analysis were assigned as male ($n = 317$) or female avatars ($n = 354$). Figure 1 shows a screenshot of the South Asian female instructor delivering the lecture with a Zoom interface overlaid.

¹ <https://www.reallusion.com/character-creator/>

After watching a video, an attention check was administered through a brief test to ensure that the participant had watched the video carefully. The researchers made sure that these questions include generic details and did not prime the participants in any way. An example of an attention test question is “What was the color of the shirt that the lecturer was wearing?” Data from the participants who could not answer at least two of the four questions correctly was excluded from the analysis. After the attention test, participants were provided with different questions for teaching evaluation.

2.3. Materials

Video vignettes and follow up questions were utilized to collect data via Qualtrics. We administered “Facilitating Learning” and “Engaging” subscales of Pedagogical Agent Persona questionnaire (Ryu and Baylor, 2005). Inspired from similar previous research (Basow, 1995; Boring et al., 2016), each item was treated as an independent variable.

2.3.1. Facilitating learning

The subscale consisted of 10 items and responses ranging on a 5-point Likert Scale, ranging from 1 = strongly disagree to 5 = strongly agree. Higher score on each item represented a more positive evaluation of the online teaching agent and greater perceived learning. Example of items are “The agent kept my attention” and “The agent encouraged me to reflect what I was learning.” The Cronbach alpha was 0.94, indicating strong

internal reliability. Please refer to Table 3 to view each item of this subscale.

2.3.2. Engaging

The subscale consisted of five items and responses ranging on a 5-point Likert Scale, ranging from 1 = strongly disagree to 5 = strongly agree. Higher score indicated that participants found the agent to be more engaging. Samples items are “The agent was entertaining” and “The agent was friendly.” The Cronbach alpha was 0.84, indicating strong internal reliability. Please refer to Table 3 to view each item of this subscale.

2.4. Analytical approach

To compare students’ evaluation of female and male instructors teaching, several independent-samples *t*-test were conducted. Thus, fifteen two-tailed independent *t*-tests were administered to explore whether instructor’s gender contributed to participants’ teaching evaluation for each of the items in subscales “facilitating learning” and “engagement.” Benjamini-Hochberg correction was used to adjust the *p*-values to avoid inflated type 1 error rate (see Supplementary material). Given we are exploring gender bias in a new context, we used a false discovery rate of 0.2. It is important to note, findings based on mean scores have also been reported to offer both perspectives to readers.

Exploratory factor analysis was also administered to observe whether 15 items from the two subscales potentially tapped clearly into different clusters. Barring one item, “instructor was enjoyable,” all the other items clearly clustered into the two original subscales. For that particular item, since factor loadings were higher for its original subscale (facilitating learning), it was considered to be part of factor 1. Therefore, original subscales were used when differences were explored based on mean scores. Refer to Table 1 to view factorial structure and factor loadings for the two factors.

All measures and conditions have been reported above and data analysis was administered after data collection ended. The data is not available on an online repository due to lack of approval by the ethics committee to share deidentified data due to the sensitive nature of the study regarding student biases.

3. Results

Table 2 shows descriptive statistics of the participants for each of the condition (Male Participant-Male Avatar, Female Participant-Male Avatar, Male Participant-Female Avatar, and Female Participant-Female Avatar). Mean scores suggest participants largely had a stronger preference for male teaching instructors compared to female instructors. Participant gender has not been used as a moderator in the analyses presented below.

We anticipated that participants would show greater bias against female virtual instructors compared to male instructors for variables related to facilitating learning and engagement. Potential gender bias was explored by administering multiple independent *t*-tests (see Table 3). In relation to facilitating learning, participants felt that male instructors made the instructions more interesting

TABLE 1 Factorial structure of “facilitating learning” and “engaging” subscales.

Items	Factor loadings	
	Factor 1	Factor 2
The agent led me to think more deeply about the presentation	0.720	
The agent made the instruction interesting	0.726	
The agent encouraged me to reflect what I was learning	0.697	
The agent kept my attention	0.759	
The agent presented the material effectively	0.801	
The agent helped me to concentrate on the presentation	0.738	
The agent focused me on the relevant information	0.778	
The agent improved my knowledge of the content	0.804	
The agent was interesting	0.688	
The agent was enjoyable	0.632	0.553
The agent was expressive		0.826
The agent was enthusiastic		0.844
The agent was entertaining		0.807
The agent was motivating		0.729
The agent was friendly		0.547

Rotated component matrix for “facilitating learning” and “engagement” subscales from Pedagogical Agent Persona questionnaire (Ryu and Baylor, 2005); extraction method was principal component analysis, and rotation method was Varimax.

[$t_{(666)} = -2.22, p = 0.027, d = -0.17$], kept a stronger hold on their attention [$t_{(666)} = -2.21, p = 0.027, d = -0.17$], and were more interesting themselves [$t_{(664)} = -2.84, p = 0.004, d = -0.22$], compared to female teaching instructors. Non-significant findings suggest that the data trends toward participants feeling that male teaching instructors made them reflect more on their learning ($p = 0.08$) and that they were also more enjoyable ($p = 0.069$), compared to female instructors.

Variables related to levels of engagement with the teaching instructors suggests that male instructors were perceived to be more expressive [$t_{(665)} = -2.50, p = 0.012, d = -0.19$], enthusiastic [$t_{(662)} = -2.89, p = 0.004, d = -0.22$], and entertaining [$t_{(663)} = -2.05, p = 0.040, d = -0.16$], compared to female teaching instructors. Non-significant findings suggest male instructors trend toward being more motivating compared to female instructors [$t_{(663)} = -1.91, p = 0.056$].

Independent *t*-test based on mean score of subscale, facilitating learning, also trends toward suggesting that male avatar teaching instructors ($M = 4.79, SD = 1.22$) were perceived to have facilitated greater learning compared to female avatar teaching instructors [$M = 4.62, SD = 1.79$], [$t_{(666)} = -1.79, p = 0.073, d = -0.13$]. Independent *t*-test based on total score of subscale, engagement, shows that male teaching instructors ($M = 4.28, SD = 1.28$) were perceived to be more engaging by the participants compared to

TABLE 2 Mean and SD of male and female participants for male and female avatar teaching instructors.

Variable	Male instructor				Female instructor			
	Male participants (N = 69)		Female participants (N = 240)		Male participants (N = 89)		Female participants (N = 258)	
	M	SD	M	SD	M	SD	M	SD
The agent led me to think more deeply about the presentation	4.54	1.41	4.59	1.53	4.54	1.59	4.43	1.62
The agent made the instruction interesting	4.74	1.46	4.71	1.54	4.58	1.52	4.41	1.67
The agent encouraged me to reflect what I was learning	4.75	1.39	4.78	1.53	4.69	1.52	4.51	1.58
The agent kept my attention	4.8	1.43	4.37	1.75	4.24	1.72	4.11	1.8
The agent presented the material effectively	5.22	1.24	5.26	1.41	5.08	1.4	5.24	1.35
The agent helped me to concentrate on the presentation	4.39	1.61	4.45	1.65	4.33	1.66	4.28	1.7
The agent focused me on the relevant information	5.12	1.32	5.2	1.41	5	1.42	5.22	1.4
The agent improved my knowledge of the content	5.45	1.23	5.34	1.4	5.26	1.35	5.2	1.51
The agent was interesting	4.86	1.46	4.77	1.58	4.43	1.74	4.41	1.7
The agent was enjoyable	4.61	1.45	4.57	1.57	4.37	1.73	4.35	1.67
The agent was expressive	4.03	1.68	3.75	1.68	3.7	1.64	3.43	1.76
The agent was enthusiastic	4.55	1.56	4.27	1.6	4.11	1.6	3.91	1.71
The agent was entertaining	4.12	1.63	3.92	1.6	3.9	1.62	3.64	1.68
The agent was motivating	4.23	1.48	4.25	1.55	4.08	1.71	3.97	1.67
The agent was friendly	5.09	1.48	5.15	1.35	4.9	1.55	4.96	1.4

female teaching instructors [M = 4.00, SD = 1.35, $t_{(665)} = -2.74$, $p = 0.006$, $d = -0.21$]. Please view the [Supplementary material](#) to explore the analyses after adjusting for multiple comparisons using the Benjamini-Hochberg correction. “The agent made the instruction interesting,” “the agent kept my attention,” and “the agent was entertaining,” became non-significant post Benjamini-Hochberg correction, largely indicating that participants held stronger biases against female lecturers in perceiving how engaging they were compared to male lecturers. However, owing to the exploratory nature of the study and analysis we have discussed the findings in the discussion section according to the original analysis.

It is also important to note that no significant differences were found between South Asian and White female lecturers for all items in subscales “facilitating learning” and “engaging.” In terms of male lecturers, it was found that South Asian lecturers were perceived to have made instructions more interesting [$t_{(314)} = 2.35$, $p = 0.019$, $d = 0.265$], to be more expressive [$t_{(314)} = 2.01$, $p = 0.045$, $d = 0.22$], and be more enthusiastic [$t_{(304)} = 2.30$, $p = 0.022$, $d = 0.26$] compared to White lecturers. No significant differences were found for other items in subscales “facilitating learning” and “engaging.”

4. Discussion

The current study explored gender bias against women in higher education in the UAE, primarily in the context of South

Asian and White instructors, ethnicities that represent two of the largest proportions of the expatriate population in the region². We examined whether perception of male and female teaching agents differed in terms of how they facilitate learning and level of engagement, using an experimental design. Findings suggest that largely participants had stronger preferences toward male instructors compared to female instructors. These differences emerged strongly in variables which represent personality characteristics, such as enthusiasm and how interesting they are perceived as, compared to variables which represent teaching and presentation, such as effective presentation and improvement of participants’ knowledge.

Male instructors were perceived to be more entertaining, enthusiastic, and expressive than female instructors. This finding is in line with previous research suggesting that competent lecturers are normally described as charismatic, fascinating and charming vs. boring with good interpersonal skills (Di Battista et al., 2020). Drawing from the shifting standard literature (Biernat and Manis, 1994), which suggests that stereotyped expectations lead to different evaluative standards, it could be argued that participants had lower expectations from male lecturers in terms of expression, entertainment and enthusiasm compared to female lecturers.

2 <https://www.globalmediainsight.com/blog/uae-population-statistics/>

TABLE 3 Independent sample *t*-tests for the questionnaire items by instructor gender (female and male).

	Female teaching instructor (N = 354)		Male teaching instructor (N = 317)		t-test		d
	Mean	SD	Mean	SD	t	p	
Facilitating learning							
The agent led me to think more deeply about the presentation	4.46	1.60	4.55	1.51	-0.77	0.440	-0.06
The agent made the instruction interesting	4.44	1.63	4.72	1.51	-2.22	0.027	-0.17
The agent encouraged me to reflect what I was learning	4.55	1.57	5.76	1.50	-1.71	0.087	-0.13
The agent kept my attention	4.14	1.78	4.44	1.69	-2.21	0.027	-0.17
The agent presented the material effectively	5.20	1.36	5.24	1.38	-0.38	0.702	-0.03
The agent helped me to concentrate on the presentation	4.28	1.70	4.43	1.62	-1.18	0.236	-0.09
The agent focused me on the relevant information	5.16	1.40	5.16	1.39	-0.01	0.989	-0.00
The agent improved my knowledge of the content	5.21	1.48	5.34	1.37	-1.18	0.238	-0.09
The agent was interesting	4.41	1.72	4.77	1.56	-2.84	0.004	-0.22
The agent was enjoyable	4.35	1.69	4.58	1.54	-1.81	0.069	-0.14
Engaging							
The agent was expressive	3.48	1.73	3.81	1.67	-2.50	0.012	-0.19
The agent was enthusiastic	3.95	1.68	4.32	1.58	-2.89	0.004	-0.22
The agent was entertaining	3.69	1.67	3.95	1.59	-2.05	0.040	-0.16
The agent was motivating	3.98	1.67	4.22	1.52	-1.91	0.056	-0.14
The agent was friendly	4.94	1.45	5.12	1.37	-1.64	0.101	-0.12

Bold values mean *p* < 0.05, italic values mean that *p* is marginally significant.

This could lead to females being judged more harshly on such parameters.

Enthusiasm is seen as a component of high-quality instruction, which seems to have a positive effect on learners’ engagement, motivation and willingness to learn (Turner et al., 1998; Witcher and Onwuegbuzie, 1999). Personality variables such as charisma have been shown to positively influence evaluation of instructors despite lack of expertise in the topic and conflicting information was delivered (Naftulin et al., 1973). Previous research suggests that lecturers’ attitudes and behaviors may influence students’ motivation (Misbah et al., 2015). It is believed that when lecturers demonstrate positive attitudes, students may become more motivated to learn. This is in line with the significant trend suggesting that male online agents were perceived to be more motivating than female online agents.

These findings are concerning as lecturers’ characteristics that are believed to influence students, such as communicational behavior, different interpersonal qualities such as enthusiasm and expressiveness, among others (Doménech-Betoret and Gómez-Artiga, 2014), were ranked higher in male instructors in comparison to female instructors. In fact, these characteristics are

often included in teaching evaluations which further influence progressions and salary increments in academia (Basow, 1995; Spooren et al., 2013).

Thus, display of similar emotions in men and women can be interpreted differently by students and, as a result, lead to different consequences when it comes to students’ evaluation of teaching. Such findings carry over in perceived leadership traits in men and women. In line with current findings, where female instructors are perceived to be less expressive and engaging, previous literature also seems to suggest that minor or moderate emotions displayed by female leaders are perceived negatively (Brescoll, 2016).

While evaluating the degree to which instructors were able to facilitate learning, participants have demonstrated significant biases against female instructors in several domains. For instance, students think that male lecturers made instructions more interesting and were able to keep students’ attention better. Instructors themselves were also seen as being more interesting that goes in line with students’ evaluation of how engaging the lecturers are. Male instructors were also trending toward being considered more enjoyable and contributing toward greater reflection on learning. It is possible that such biases contribute to putting male

lecturers in higher regard, leading to them being called “professors” compared to female lecturers being referred to as “teachers” (Miller and Chamberlin, 2000; Renström et al., 2021). Moreover, evaluation of objective student learning was not possible since participants only watched a short video. However, in studies, where student learning was compared, no differences were observed between female and male lecturers (Boring, 2017). Perhaps this could be attributed to perceived personality characteristics of male lecturers being viewed more favorably compared to female lecturers. Thus, it seems a plausible conclusion that the participants’ biases should explain the emerging differences. Also, while 74% of the participants are female, male instructors were consistently preferred in all domains of teaching evaluation. This finding is not in line with previous research which suggests that typically students favor instructors of the same gender as themselves [review in Kim and Baylor, 2016]. Future research could moderate the analysis using participant gender to further investigate lack of in-group preference.

These differences in rating could lead to significant differences not only in students’ learning experience but also in their eagerness to take a particular course depending on the gender of the instructor. The latter was supported by Arrona-Palacios et al. (2020), who found that students tended to favor their male professors over female professors, recommending them more. This is concerning because many universities are offering optional modules, which might be withdrawn if students systematically opt out, which might seriously impact female professors’ academic careers if they are teaching such modules.

Post Benjamini-Hochberg correction, findings largely indicate that participants held stronger biases against female lecturers in perceiving how engaging they were compared to male lecturers. It was found that in some domains, gender did not significantly influence teaching evaluation. Instructors were relatively similarly rated despite their gender when participants evaluated professionalism, presentation skills, or knowledge, based on responses to questions related to improvement of knowledge, focus on relevant information, and encouragement to think more deeply. While both genders were given high ratings, which is a positive sign, it is important to note that despite no statistically significant differences, ratings consistently showed preference for male instructors over female instructors. Future research should investigate this further as literature in this domain is divided. Current findings are not in line with some studies suggesting that participants felt that male instructors possessed greater expertise in subject matter compared to female instructors (Miller and Chamberlin, 2000; Renström et al., 2021). However, findings support a few previous studies (Boyd and Grant, 2005; Zikhali and Maphosa, 2012), suggesting that competence was not gender-determined. Further research should look more deeply into ratings which were found only marginally significant, such as how enjoyable and motivating was instructor and if the instructor encouraged students to reflect on what they learned. Researchers could also explore similar domains of teaching evaluation in live teaching in the region and compare findings to studies representing an online teaching platform.

Given the importance of ethnicity for teaching evaluation highlighted by previous research (Green et al., 2012; Baker et al., 2013; Chávez and Mitchell, 2020) and racial representation in the

region, developed videos displayed White and South Asian male and female instructors. As mentioned earlier, adding both races (White and South Asian) provided a balanced representation of ethnicities in the region. However, it is important to note that while female lecturers did not differ in how they were perceived based on their race, for male avatars, South Asian lecturers were perceived to be more interesting, expressive, and enthusiastic compared to White lecturers. This finding is similar to previous research which found differences in evaluation based on race in the context of higher education (Smith, 2007; Chisadza et al., 2019). Therefore, it is possible that the evidence of gender bias in these variables was facilitated by South Asian male lecturers receiving higher scores and increasing the average scores males received, compared to female lecturers. However, race in isolation did not significantly contribute to overall finding of gender bias, as it did not contribute to any differences in the female category. Future research could explore the role of race as well in the context of higher education in the UAE.

Additionally, Observed differences cannot be attributed to differences in performance because each avatar had the exact same script, nor can it be attributed to the differences in personal presentations, as all avatars wore the same clothes, had the same background, and held the same expressions. Thus, it is possible that emerged differences are explained by the participants’ biases and shifting standards model leading to gender bias. However, to explore the possible cultural difference, future research could look at the gender equity value assigned to national origins as an independent predictor of the outcomes by gender.

This study provides several contributions to the emerging literature about online teaching and gender biases that may impact teaching evaluation. To our knowledge it is the only study exploring gender biases based on an online teaching platform (Zoom, in this case) using avatars representing instructors. It is also the first study exploring gender bias in academia in the UAE. Studying underrepresented samples and inclusion of South Asian virtual instructors are the key strengths of the study. The research also utilized nuanced variables such as facilitating learning and engagement, compared to a general teaching evaluation questionnaire. We found that female instructors are discriminated against in this online format, similar to the face-to-face classes. This confirms that gender bias continues to exist in academia, especially evident in terms of personality traits. It is expected that these evaluations are likely to affect how women are perceived by colleagues, supervisors and management. Thus, our findings raise concerns regarding progression of women in academic careers.

In terms of limitations, recruiting a more diverse sample would have enhanced existing findings. Future research could aim for a more balanced representation of participants. Having avatars instead of real people might also be seen as a limitation, but since the results were in line with previous studies, a controlled experimental design may also be viewed as an appropriate scenario for addressing these biases. Another limitation of the study is having relatively young avatars, which should be addressed by future research comparing teaching evaluations for older and younger instructors. Statistically, multiple *t*-tests may have raised type 1 error rate and the effect sizes are relatively low, therefore results should be interpreted with caution. It is also important to diversify the context of the research by including teaching

videos in other fields, including STEM subjects since the choice of subject might also have a significant impact on the ratings. Thus, future studies could look at “subject” as a factor, keeping everything constant and manipulating the content itself to see if differences arise based on instructor gender and the subject being taught.

It could be concluded that overall, female lecturers continue to face discriminatory attitudes in higher education. These attitudes are measured by student evaluations, that are widely used in academia, and that have typically produced a disadvantage for female instructors. Previous research suggests that there is no evidence that this is an exception rather than the rule (Boring et al., 2016; Mengel et al., 2017). Thus, taking into account the biased nature of teaching evaluations, these should be used with caution as relying on them could be discriminatory for female lecturers. In fact, other measures such as open forums and discussion may make students more aware of their potential biases. However, previous research highlighted this problem doesn't exist only at the students' level, there is a relative reluctance amongst male faculty within science, technology, engineering, and mathematics (STEM) to accept the evidence of gender-bias research (Handley et al., 2015). Therefore, more complex measurements are required to further understand and address gender discrimination.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Middlesex University Dubai. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

References

- Arrona-Palacios, A., Okoye, K., Camacho-Zuñiga, C., Hammout, N., Luttmann-Nakamura, E., Hosseini, S., et al. (2020). Does professors' gender impact how students evaluate their teaching and the recommendations for the best professor? *Heliyon* 6:e05313. doi: 10.1016/j.heliyon.2020.e05313
- Babin, J. J., and Hussey, A. (2023). Gender penalties and solidarity – teaching evaluation differentials in and out of stem. *Econ. Lett.* 226:111083. doi: 10.1016/j.econlet.2023.111083
- Babin, J. J., Hussey, A., Nikolsko-Rzhevskyy, A., and Taylor, D. A. (2020). Beauty premiums among academics. *Econ. Educ. Rev.* 78:102019. doi: 10.1016/j.econedurev.2020.102019
- Baker, B. D., Oluwole, J., and Green, P. (2013). The legal consequences of mandating high stakes decisions based on low quality information: teacher evaluation in the race-to-the-top era. *Educ. Eval. Policy Anal. Arch.* 21, 1–71. doi: 10.14507/epaa.v21n5.2013
- Bao, W. (2020). COVID-19 and online teaching in higher education: a case study of Peking University. *Hum. Behav. Emerg. Technol.* 2, 113–115. doi: 10.1002/hbe.2191
- Basow, S. A. (1995). Student evaluations of college professors: when gender matters. *J. Educ. Psychol.* 87, 656–665. doi: 10.1037/0022-0663.87.4.656
- Baylor, A. L. (2005). “The impact of pedagogical agent image on affective outcomes,” in *International Conference on Intelligent User Interfaces* (San Diego, CA), 29.
- Baylor, A. L., and Kim, Y. (2004). “Pedagogical agent design: the impact of agent realism, gender, ethnicity, and instructional role,” in *International Conference on Intelligent Tutoring Systems* (Springer), 592–603. doi: 10.1007/978-3-540-30139-4_56
- Baylor, A. L., and Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *Int. J. Artif. Intell. Educ.* 15, 95–115.
- Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: evidence relating to the question of gender bias in teaching evaluation. *J. Educ. Psychol.* 74:170. doi: 10.1037/0022-0663.74.2.170
- Berezvai, Z., Lukáts, G. D., and Molontay, R. (2021). Can professors buy better evaluation with lenient grading? The effect of grade inflation on student evaluation of teaching. *Assess. Eval. High. Educ.* 46, 793–808. doi: 10.1080/02602938.2020.1821866
- Biernat, M., and Manis, M. (1994). Shifting standards and stereotype-based judgments. *J. Pers. Soc. Psychol.* 66:5. doi: 10.1037/0022-3514.66.1.5

Author contributions

OK, NL, and SK contributed to conception and design of the study. NL performed the statistical analysis. All authors contributed toward the conception and design of the study, data collection, manuscript revision, read, and approved the submitted version.

Acknowledgments

The authors would like to thank Zainab Mohammad Ali, Emmeline Pereira, and Natasha Azoury for their assistance with data collection.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2023.1158132/full#supplementary-material>

- Boring, A. (2017). Gender biases in student evaluations of teaching. *J. Publ. Econ.* 145, 27–41. doi: 10.1016/j.jpubecon.2016.11.006
- Boring, A., Ottoboni, K., and Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Res.* 1–11. doi: 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1
- Boyd, E., and Grant, T. (2005). Is gender a factor in perceived prison officer competence? Male prisoners' perceptions in an English dispersal prison. *Crim. Behav. Ment. Health* 15, 65–74. doi: 10.1002/cbm.37
- Brescoll, V. L. (2016). Leading with their hearts? How gender stereotypes of emotion lead to biased evaluations of female leaders. *Leadersh. Q.* 27, 415–428. doi: 10.1016/j.leaqua.2016.02.005
- Ceci, S. J., and Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3157–3162. doi: 10.1073/pnas.1014871108
- Centra, J. A., and Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *J. High. Educ.* 71, 17–33. doi: 10.1080/00221546.2000.11780814
- Chávez, K., and Mitchell, K. M. (2020). Exploring bias in student evaluations: gender, race, and ethnicity. *Polit. Sci. Polit.* 53, 270–274. doi: 10.1017/S1049096519001744
- Chisadza, C., Nicholls, N., and Yitbarek, E. (2019). Race and gender biases in student evaluations of teachers. *Econ. Lett.* 179, 66–71. doi: 10.1016/j.econlet.2019.03.022
- Di Battista, S., Pivetti, M., and Berti, C. (2020). Competence and benevolence as dimensions of trust: lecturers' trustworthiness in the words of Italian students. *Behav. Sci.* 10:143. doi: 10.3390/bs10090143
- Doménech-Betoret, F., and Gómez-Artiga, A. (2014). The relationship among students' and teachers' thinking styles, psychological needs and motivation. *Learn. Individ. Differ.* 29, 89–97. doi: 10.1016/j.lindif.2013.10.002
- Dziuban, C., and Moskal, P. (2011). A course is a course is a course: factor invariance in student evaluation of online, blended and face-to-face learning environments. *Intern. High. Educ.* 14, 236–241. doi: 10.1016/j.iheduc.2011.05.003
- Fan, Y., Shepherd, L. J., Slavich, E., Waters, D. D., Waters, D., Waters, D., et al. (2019). Gender and cultural bias in student evaluations: why representation matters. *PLOS ONE* 14:209749. doi: 10.1371/journal.pone.0209749
- Feldman, K. A. (1993). College students' views of male and female college teachers: part ii—evidence from students' evaluations of their classroom teachers. *Res. High. Educ.* 34, 151–211. doi: 10.1007/BF00992161
- Fuegen, K., Biernat, M., Haines, E., and Deaux, K. (2004). Mothers and fathers in the workplace: how gender and parental status influence judgments of job-related competence. *J. Soc. Issues* 60, 737–754. doi: 10.1111/j.0022-4537.2004.00383.x
- Gillies, D. (2008). Student perspectives on videoconferencing in teacher education at a distance. *Dist. Educ.* 29, 107–118. doi: 10.1080/01587910802004878
- Green, P. C. III, Baker, B. D., and Oluwole, J. (2012). *The Legal and Policy Implications of Value-Added Teacher Assessment Policies*. BYU Educ. & LJ. 1.
- Hamermesh, D. S., and Parker, A. M. (2005). Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity. *Econ. Educ. Rev.* 24, 369–376. doi: 10.1016/j.econedurev.2004.07.013
- Handley, I. M., Brown, E. R., Moss-Racusin, C. A., and Smith, J. L. (2015). Quality of evidence revealing subtle gender biases in science is in the eye of the beholder. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13201–13206. doi: 10.1073/pnas.1510649112
- Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Educ.* 4:1304016. doi: 10.1080/2331186X.2017.1304016
- Kim, Y., and Baylor, A. L. (2016). Research-based design of pedagogical agent roles: a review, progress, and recommendations. *Int. J. Artif. Intell. Educ.* 26, 160–169. doi: 10.1007/s40593-015-0055-y
- Kim, Y., and Wei, Q. (2011). The impact of learner attributes and learner choice in an agent-based environment. *Comput. Educ.* 56, 505–514. doi: 10.1016/j.compedu.2010.09.016
- Klein, U. (2016). Gender equality and diversity politics in higher education: conflicts, challenges and requirements for collaboration. *Women's Stud. Int. Forum* 54, 147–156. doi: 10.1016/j.wsif.2015.06.017
- Kreitzer, R. J., and Sweet-Cushman, J. (2021). Evaluating student evaluations of teaching: a review of measurement and equity bias in SETs and recommendations for ethical reform. *J. Acad. Ethics* 20, 73–84. doi: 10.1007/s10805-021-09400-w
- MacNell, L., Driscoll, A., and Hunt, A. N. (2015). What's in a name: exposing gender bias in student ratings of teaching. *Innov. High. Educ.* 40, 291–303. doi: 10.1007/s10755-014-9313-4
- Mama, A. (2003). Restore, Reform but do not transform: the gender politics of higher education in Africa. *J. High. Educ. Africa* 1, 101–125. doi: 10.57054/jhea.v1i1.1692
- Marsh, H. W. (1980). The influence of student, course, and instructor characteristics in evaluations of university teaching. *Am. Educ. Res. J.* 17, 219–237. doi: 10.3102/00028312017002219
- Mengel, F., Sauermann, J., and Zöltz, U. (2017). Gender bias in teaching evaluations. *J. Eur. Econ. Assoc.* 17, 535–566. doi: 10.2139/ssrn.3037907
- Miller, J., and Chamberlin, M. (2000). Women are teachers, men are professors: a study of student perceptions. *Teach. Sociol.* 28, 283–298. doi: 10.2307/1318580
- Misbah, Z., Gulikers, J., Maulana, R., and Mulder, M. (2015). Teacher interpersonal behaviour and student motivation in competence-based vocational education: evidence from Indonesia. *Teach. Teach. Educ.* 50, 79–89. doi: 10.1016/j.tate.2015.04.007
- Naftulin, D. H., Ware, J. E. Jr., and Donnelly, F. A. (1973). The doctor fox lecture: a paradigm of educational seduction. *Acad. Med.* 48, 630–635. doi: 10.1097/00001888-197307000-00003
- O'Connor, P. (2020). Why is it so difficult to reduce gender inequality in male-dominated higher educational organizations? A feminist institutional perspective. *Interdiscip. Sci. Rev.* 45, 207–228. doi: 10.1080/03080188.2020.1737903
- Ovseiko, P. V., Chapple, A., Edmunds, L. D., and Ziebland, S. (2017). Advancing gender equality through the Athena SWAN Charter for Women in Science: an exploratory study of women's and men's perceptions. *Health Res. Policy Syst.* 15:12. doi: 10.1186/s12961-017-0177-9
- Park, S. (2020). Seeking changes in ivory towers: the impact of gender quotas on female academics in higher education. *Women's Stud. Int. Forum* 79:102346. doi: 10.1016/j.wsif.2020.102346
- Rapanta, C., Botturi, L., Goodyear, P., Guàrdia, L., and Koole, M. (2020). Online university teaching during and after the COVID-19 crisis: refocusing teacher presence and learning activity. *Postdigit. Sci. Educ.* 1, 132–145. doi: 10.1007/s42438-020-00155-y
- Reid, L. D. (2010). The role of perceived race and gender in the evaluation of college teaching on RateMyProfessors.Com. *J. Divers. High. Educ.* 3, 137–152. doi: 10.1037/a0019865
- Renström, E. A., Sendén, M. G., and Lindqvist, A. (2021). Gender stereotypes in student evaluations of teaching. *Front. Educ.* 5:571287. doi: 10.3389/feduc.2020.571287
- Rudman, L. A., and Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *J. Soc. Issues* 57, 743–762. doi: 10.1111/0022-4537.00239
- Ryu, J., and Baylor, A. L. (2005). The psychometric structure of pedagogical agent persona. *Technol. Instruct. Cogn. Learn.* 2:291.
- Sinclair, L., and Kunda, Z. (2000). Motivated stereotyping of women: she's fine if she praised me but incompetent if she criticized me. *Pers. Soc. Psychol. Bull.* 26, 1329–1342. doi: 10.1177/0146167200263002
- Smith, B. P. (2007). Student ratings of teaching effectiveness: an analysis of end-of-course faculty evaluations. *Coll. Stud. J.* 41, 788–800.
- Spooren, P., Brockx, B., and Mortelmans, D. (2013). On the validity of student evaluation of teaching: the state of the art. *Rev. Educ. Res.* 83, 598–642. doi: 10.3102/0034654313496870
- Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: a theoretical and empirical analysis. *Basic Appl. Soc. Psychol.* 42, 276–294. doi: 10.1080/01973533.2020.1756817
- Tatro, C. N. (1995). Gender effects on student evaluations of faculty. *J. Res. Dev. Educ.* 28, 169–173.
- Turner, J. C., Meyer, D. K., Cox, K. E., Logan, C., DiCintio, M., and Thomas, C. T. (1998). Creating contexts for involvement in mathematics. *J. Educ. Psychol.* 90, 730–745. doi: 10.1037/0022-0663.90.4.730
- Wachtel, H. K. (1998). Student evaluation of college teaching effectiveness: a brief review. *Assess. Eval. High. Educ.* 23, 191–212. doi: 10.1080/0260293980230207
- Williams, W. M., and Ceci, S. J. (2015). National hiring experiments reveal 2: 1 faculty preference for women on stem tenure track. *Proc. Natl. Acad. Sci. U.S.A.* 112, 5360–5365. doi: 10.1073/pnas.1418878112
- Witcher, A., and Onwuegbuzie, A. J. (1999). "Characteristics of effective teachers: perceptions of preservice teachers," in *Annual Meeting of the Mid-South Educational Research Association (MSERA)* (Point Clear, AL). Available online at: <https://files.eric.ed.gov/fulltext/ED438246.pdf>
- Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teach. High. Educ.* 12, 55–76. doi: 10.1080/13562510601102131
- Zikhali, J., and Maphosa, C. (2012). Exploring college students' attitudes towards female lecturers' competence in teacher education colleges in Masvingo Province, Zimbabwe. *Anthropologist* 14, 393–399. doi: 10.1080/09720073.2012.11891262