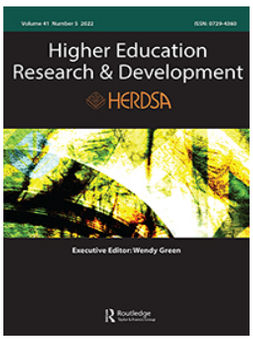

Higher Education Research & Development

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cher20

Student evaluation of teaching: gender bias in a country at the forefront of gender equality

Margret Sigrun Sigudardottir, Gudbjorg Linda Rafnsdottir, Anna Helga Jónsdóttir & Dadi Mar Kristofersson

To cite this article: Margret Sigrun Sigudardottir, Gudbjorg Linda Rafnsdottir, Anna Helga Jónsdóttir & Dadi Mar Kristofersson (2022): Student evaluation of teaching: gender bias in a country at the forefront of gender equality, Higher Education Research & Development, DOI: 10.1080/07294360.2022.2087604

To link to this article: https://doi.org/10.1080/07294360.2022.2087604

HIGHER EDUCATION RESEARCH & DEVELOPMENT
https://doi.org/10.1080/07294360.2022.2087604

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

Student evaluation of teaching: gender bias in a country at the forefront of gender equality

Margret Sigrun Sigudardottir [a], Gudbjorg Linda Rafnsdottir [b], Anna Helga Jónsdóttir [c] and Dadi Mar Kristofersson [d]

[a]Faculty of Business Administration, University of Iceland, Reykjavik, Iceland; [b]Faculty of Social and Human Sciences, University of Iceland, Reykjavik, Iceland; [c]Faculty of Physical Sciences, University of Iceland, Reykjavik, Iceland; [d]Faculty of Economics, University of Iceland, Reykjavik, Iceland

ABSTRACT

Research on gender equality in Higher Education indicates that female academics might not have the same opportunities for promotion as male academics. One of the are as in Higher Education where gender bias has been reported is in Student Evaluations of Teaching (SET). The aim of this article is to analyse possible gendered characteristics of communication between students and teachers in higher education, as reflected in SET. The article builds on mixed-method research. Five years of SET responses were analysed quantitatively and one year of open answers was analysed qualitatively. The qualitative data was then coded quantitatively. Iceland offers an excellent case for this study as it has been a forerunner in gender equality for years. The quantitative analysis shows that male students rate female teachers lower than their male counterparts, and the qualitative analyses indicate that the evaluation pattern of male and female teachers differs. Additionally, differences were found in the comments received by the teachers. Comments on male teachers referred to subject knowledge, while female teachers received comments on their service to students and relatability. There thus seems to be a gender bias in SET in Iceland despite its reputation for gender equality.

ARTICLE HISTORY
Received 9 September 2021
Accepted 5 April 2022

KEYWORDS
Higher education; student evaluations of teaching (SET); gender bias; gender equality

Introduction

Research on gender equality in higher education indicates that the opportunities for promotion differ between male and female academics and that women are underrepresented as academic leaders (Cech & Blair-Loy, 2019). This is true even in the Nordic countries (Nielsen, 2017; Teigen & Skjeie, 2017), despite their well-known performance in gender equality (World Economic Forum, 2021).

The term '*leaky pipes*' is sometimes used to describe the reason for this; women 'leak out' of the 'pipes of promotion' on their way up the university hierarchy, as they face greater hindrances in their career promotion than their male counterparts (Anders, 2004; Gasser & Shaffer, 2014). Nevertheless, there is a lack of understanding of the types of hindrances involved. A recent experimental study of the evaluations of

CONTACT Margret Sigrun Sigudardottir mss@hi.is

applications for academic positions did not show bias against women in the Nordic countries (Carlson et al., 2021). However, the authors point out the importance of analysing possible gender bias after people are hired in academia. In a recent article, Lipton (2021), drawing upon data from Australian universities, showed double standards imposed upon academic women, where they on the one hand are pressed to adhere to the 'empowered woman trope' but on the other to dress 'properly'. Lipton argues that this reflects the reality that: 'Academic women are not simply judged on their meritocratic performance as scholars but are judged by the equally subjective standard of how well they fit into the masculine culture of the contemporary university' (2021, p. 768). Lipton (2021) further points out that in academia, appearance is most often cited in relation to Student Evaluations of Teaching (SET). A literature review conducted by Heffernan (2021) further indicates that SET outcomes are influenced by racist, sexist, and homophobic prejudices. Research on SET also indicates that the students' answers are based on how well they like the teacher rather than the teaching, thus indicating that if the student has a preference for either male or female stereotypes, this can affect the overall rating of the courses taught (Clayson, 2022). Concerns about the validity of SET are not new, as Kulik's overview from 2001 outlines, but are still a relevant study due to their prominence in the evaluation of teaching in higher education.

We agree with Lipton (2021) that gender is partly performatively produced. Therefore, we also argue that possible gendered outcomes of SET can be meaningful for gender equality in higher education and the leaky pipelines phenomenon (Anders, 2004; Gasser & Shaffer, 2014), promotion and tenure decisions (Algozzine et al., 2004). The aim of the article is to examine possible gender differences as reflected in SET at the University of Iceland (UI), which is by far the largest university in Iceland. The university was a part of Carlson et al.'s (2021) Nordic study showing no bias against women in the academic recruitment process. As Iceland has been the frontrunner of gender equality in the world since 2009 (World Economic Forum, 2021), the study analyses if and how this strong position of gender equality in the country is possibly reflected in the outcome of student's answers in SET. More concretely, we ask whether we find gendered characteristics of communication between teachers and students in higher education, as reflected in SET. Given the status of Iceland as a leader in gender equality (World Economic Forum, 2021), it would be fair to expect negligible gender bias in our study. If that is the case, it gives indications that general gender equality in society reduces gender bias in SET. However, as Heijstra et al. (2017) found some gender bias in higher education in the country, this might not be the case.

The Icelandic setting

The University of Iceland is a Nordic public university, covering most traditional disciplines at the undergraduate and graduate level. The university has set itself the goal to become among the top universities in the world, ranked by Times Higher Education, and Shanghai Jiao Tong University list, etc. To reach that goal, research-related activities have been prioritized in the academic incentive system and in relation to promotion and permanent employment. How people have performed in teaching, for instance according to SET, is also considered.

In addition to being a broad public university with strong international connections, the university is placed in a country with a strong gender equality agenda. Iceland is number one in gender equality according to the Global Gender Gap Index (World Economic Forum, 2021), benchmarking national gender gaps based on educational, economic, political, and health-based criteria. In 2017, the country was also number one in the Economist glass-ceiling index (Economist, n.d.) in the category 'Environment for working women'.

Iceland also pops up in the international media as the cradle of gender equality in the world and a role model (Hertz, 2016; Kilpatrick, 2017; World Economic Forum, n.d.). Icelandic authorities have even turned gender equality into politics of reputation and nation branding (Einarsdóttir, 2020). This makes the UI an interesting case for possible gender disparity in SET.

Women outnumber men within UI, both as students (68%) and hired employees (55%) and women were 70% of those who graduated with PhD degrees from the university. However, only 44% of senior executives at the university are female and only 34% of full professors.

Even though women's academic career patterns and how women are addressed in SET compared with men is not determined solely by the national level of gender regime, we agree with Le Feuver (2015) that understanding the norms and expectations that surround the academics under study is important. In that respect we refer to what Thébaud and Taylor (2021) call 'The spectre of motherhood' – anticipating concerns about combining an academic career with motherhood. A study by Staub and Rafnsdóttir (2020) showed gendered societal time norms among doctorate holders in Iceland, where men felt a higher level of agency regarding work-life balance and time management than their female counterparts. Women also expressed more often than men being stressed about their often fragmented time, having to combine career and family obligations successfully. A recent mixed methods study by Staub and Heijstra (2021) further shows that male doctorate holders in Iceland, working both within and outside academia, earn significantly more than women doctorate holders. They argue that decisions made within the household, referred to as a team play, negotiation, and choices, play a defining role in the context. Even though higher education is often assumed to empower women and to diminish gender discrimination in society (Casey, 2009), these studies show that social interactions within highly educated families can affect women's position at work negatively, even in countries with a strong position of gender equality. In a similar vein, we believe that communications within universities, for example as it appears in SET, can affect academics, thus, the understanding of possible gender bias in that communications can be improved by analysing possible gendered characteristics of communication between students and teachers in higher education, as reflected in SET. Given the status of Iceland as a leader in gender equality (World Economic Forum, 2021), it would be fair to expect negligible gender bias to be found in our study.

Student evaluation of teaching (SET)

International research has indicated that there is gender discrimination in student surveys/course evaluations (Boring & Ottoboni 2016; Fan et al., 2019; Mitchell & Martin, 2018; Pounder, 2007). These findings are, nevertheless, not always significant

and could possibly be explained by differences in teaching styles between the genders (Centra & Gaubatz, 2000; Punyanunt-Carter & Carter, 2015). Although Fan et al. (2019) argue that it is hard to believe that women (and non-native English speakers) are in general much worse teachers than male native English speakers.

In an overview of the literature on SET, Pounder (2007), found gendered expectations in the way both male and female students rate male and female teachers. Based on gendered stereotypes, male teachers are perceived as more professional and female teachers as more nurturing. Further analysis of the open comments in SET indicated that women are evaluated on personality, appearance, and perception of intelligence and competency more than men. In research conducted at 21 higher education institutes in the US, Centra and Gaubatz (2000) found that students think that male teachers were better organized, while female teachers were rated higher on other aspects, such as grading and student interactions. In addition, female instructors were perceived as being more approachable than their male colleagues.

Research by Mitchell and Martin (2018) on identical online courses in political science in the US showed strong indicators of gender bias in the SET scores of male and female teachers, with men receiving higher scores from students in general. Clayson (2020) found the same pattern in different US universities. Using an analysis of students' self-reporting on how much they learned, he found that both male and female students ranked classes taught by male teachers higher than classes taught by female counterparts. This held for both business and humanities, although the preferences of the two subjects differed in terms of age and political leaning of the teacher. The aforementioned research indicates a preference for male teachers by both male and female students. According to an Australian study done by Denson et al. (2010), the gender of the student does, however, influence the evaluation of teachers, as they find that male students rate all teachers lower than female students. However, Boring (2017), found that male students tend to rate male teachers higher and female teachers lower than female students do.

The gender of teacher and student are, however, not the only variables affecting evaluations of male and female teachers in SET. In an experimental setting at a Belgian university, Hoorens et al. (2021) asked students to evaluate a hypothetical course. They found that the students' expected grade in the course influenced the grade they gave to the course, in such a way that a lower grade disadvantaged the female professors. In a survey conducted after students received grades for their courses at a Canadian university, Sinclair and Kunda (2000) found that the grade students gave to the course was more dependent on the grade the student received when the teacher was female than male; that is, a low grade from a female teacher resulted in a lower grade for the course than a low grade from a male teacher. Further, students who got a low grade from a female teacher were less satisfied with the grade than if they got a low grade from a male professor.

Peterson et al. (2019) found that informing students about potential gender bias and stereotypes improved the SET scores of female teachers by as much as half a point on a five-point scale, while it did not affect males' scores, in an experimental setting at a US university. The gender bias based on students' grades in Hoorens et al.'s (2021) study in Belgium was reduced for those students who had been assigned a self-affirmation, such as listing their values, before they answered the survey. The self-affirmation prompt did not raise the grade given to female teachers, but lowered the grade of male

teachers, leading the authors to suspect that the gender bias in SET might be due to the over-evaluation of male teachers.

Most of the research reported on was conducted in the US, although there is an indication of gender bias in research from Canada and Belgium. Given that awareness of biases (Peterson et al., 2019) and self-affirmation (Hoorens et al., 2021) can reduce bias it is interesting to see if gender bias exists in SET in a country at the very front in gender equality (World Economic Forum, 2021).

Methods

The research presented here is based on a mixed-methods approach (Creswell, 2013). That is, both qualitative and quantitative research methods are used in combination. We do this by qualitatively analysing the open answers in the SET and then quantitatively analysing the qualitative codes along with the quantitative analysis of the quantifiable answers in the survey.

Participants and data – quantitative study

The study is retrospective as all responses to the SET survey at UI, gathered over nine semesters from the 2013 fall semester to the fall semester in 2017, are analysed. There are 127,379 answers in total, given by 18,983 different students in 3521 courses with 3295 teachers. The response rate was between 35% and 53%. Most of the 3521 courses were taught more than once during the study period. A total of 70% of the students involved in the study are female and 30% male (it was not possible to register another gender in the university database during the study period). The gender proportions among the teachers are 46% female and 54%, male.

Materials – quantitative study

The SET survey at the University of Iceland covers the six areas shown in Table 1. For each of the six areas, the students specify their level of agreement with three to four statements on a five-point Likert scale, 22 statements in total. In addition, the students are asked to give the course and the teacher(s) an overall rating, both on a 1–10 scale.

Table 1. The six areas covered by the SET at the UI.

Area	Short explanation	Part
(1) *Teaching*	Factors related to the presentation of material in class and the instructor's enthusiasm for teaching	Teacher part
(2) *Academic motivation*	Covers how well the instructor encourages students to develop skills in scholarly criticism, academic independence, and critical thinking	Teacher part
(3) *Course structure*	Involves students evaluating the clarity of the course objectives and requirements as well as how achievable they are	Course part
(4) *Workload*	Covers how demanding the course is and how heavy the workload is compared to other courses that students have taken at the University of Iceland	Course part
(5) *Course outcomes*	Evaluation of the benefits of attending the course	Course part
(6) *Student contribution*	The student's own level of preparation for tackling the course subjects and the amount of work the student has put into the course	Course part

Statistical analysis – quantitative study

A mixed effects model (Pinheiro & Bates, 2006), was used to analyse the students' responses to the statements covering the six areas of the SET as well as the overall ratings of the courses and the teachers (a total of eight models).

It is important to note the fact that the 127,379 answers used for the analyses are not independent measurements since the same student answers questions for more than one course (dependency within students), and groups of students answer questions regarding the same courses and teachers (dependency within courses and teachers). In the mixed effects model setting, this is accounted for by including random effects, which in this case are random effects for students, courses, and teachers.

In addition to the random effects, the following fixed effects variables were used in all the models: gender of student, age of the student, whether the student passed the course, the average grade in the course, and the number of students registered for the course. In the models for the teachers' part (areas 1 and 2) and the overall teacher rating, the gender and age of the teachers were also included as fixed effects. In the models for the course part (areas 3–6) and the overall course rating, the following two additional variables were included: a variable stating how many teachers were involved in teaching the courses and a variable indicating if courses were taught solely by female teachers, solely by male teachers, or a mix thereof.

When using mixed effects models, it is not only possible to assess the potential differences in the fixed effects (such as gender and age) but also to quantify the different types of variation in the response variables. That is, how much of the variation is due to fixed effects, the variation between students, the variation between courses and teachers, and, finally, how big the unexplained variation in the data is.

All statistical analyses were conducted in R (R Core Team, 2020). The mixed effects models were fitted using the lme4 package (Bates et al., 2015). Due to the higher number of observations in the dataset (127,379 answers), statistical hypotheses tests have high power. This means that a small difference can result in a statistically significant result. This should be kept in mind when the results of the analyses are interpreted.

Qualitative study

The qualitative analysis was conducted on all the answers to the open questions in the school year 2016–2017. All in all, just under 4000 comments were analysed. A subsample of the comments was analysed line by line (Strauss & Corbin, 1997) in atlas.ti to create codes and themes. As the focus here is examining bias, only the codes relating to the theme of *the teacher's person* are included. These were coded as supportive or critical. Most of the comments to the open questions were very short, one sentence or even one-word comments, such as 'great' (which occurred 1674 times). All comments were analysed with logistic regression using available background information as independent variables: the gender of the student, the gender of the teacher, the age of the student and the course outcome for the student. Table 2 offers examples of comments in the codes.

Since a short sentence can express many sentiments, each sentence was often coded with multiple codes, such as the following answer coded with Organization, Explanations, and Support and Connection:

Table 2. Codes – overview and examples.

	Supportive	Critical
Explanations	The only lecturer in the course that managed to catch and keep my attention. Approached the complex subject in a humane way and explained it very well	I did not always find the teacher to be sufficiently clear in instructions about the concepts and contents of the textbook but to trust that we understood them from the book
Performance	[The teacher] is exemplary. She should advise other teachers who are not doing as well	It's obvious that [the teacher] wants to provide good teaching, but he's not succeeding
Entertaining	Very clear and entertaining teacher	The content of the course was wide and interesting, but the way it was brought to us, the way the teacher teaches, is completely boring, even if she really seems to know what she is talking about
Knowledge	Relaxed and knowledgeable in the subject he is teaching	Can be uninterested, unorganized, and does not explain very well (possibly because he knows the subject too well, I have seen this before in super intelligent specialized people)
Response	Responds quickly to mail and positively to suggestions, e.g., regarding organization in [learning management system]. Is encouraging when answering questions. Delivers assignments quickly and well. Good supervision of students considering that the course is distance learning. Encourages discussion between students	The year has been characterized by chaos and a lack of contact with the teacher, who regularly does not respond to emails and other communication
Demeanor	Very friendly	The first time I've experienced absolute disrespect of students, their knowledge and experience, and that students learn in different ways
Organization	Great, organized and fair teacher	What I feel needs to be improved in this class is either to get a new teacher or the teacher gets better organized
Approach to teaching	The teacher has a good approach to teaching, keeps the student's attention, and offers good examples that relate to the present times	I would have liked the course to be more of an interaction, more discussion with students, and also I would have liked that the teacher would make more links between [the subject] and the way society is organized now, to make the student think by themselves, instead of only receiving information and having to learn it by heart
Use of voice	Great teacher; speaks clearly and is interested in the subject	Sometimes answers arrogantly and should speak louder/clearer
Support and connection	[the teacher] makes you feel like she really cares about you learning and that you feel good. Really good classes and a good teacher. I have learned so much about life in this course	Very little help in the TAs. There was never any time for discussions

> [The teacher] covers the subject in a clear and organized way, explains the topic well, and obviously cares about her students, this shines through the teaching. She praises a lot and is encouraging.

Results

Analysis of the SET results – quantifiable answers

In the following section, the focus is on the results of the models for the overall rating of teachers and courses as the results for the teacher part (areas 1 and 2) and course part (areas 3–6), see Table 1, are similar to the overall teacher rating and course rating, respectively. A significance level of 5% is used in all statistical tests.

In the model for the overall rating of teachers, all the fixed effects are statistically significant. In addition, an interaction between the gender of the student and the gender of teachers is significant. There is no difference between how female and male students rate male teachers, but male students on average give female teachers lower grades; however, the effect is small, 0.15 on a 1–10 scale, although it needs to be considered since 80% of the data points were 7 or higher. The largest effects on the overall rating of teachers are the number of students in the course (an effect size of 0.49 when comparing courses with fewer than 25 students and courses with more than 100 students enrolled) and whether the student passed the course (students that passed the course rated teachers 0.39 points higher on an average than those that did not pass the course).

When looking at the overall rating of courses, results similar to the overall teacher rating were found. There is no difference in how female and male students rate courses solely taught by male teachers, but male students give on average lower grades to courses solely taught by female teachers. In this case, the effect is small, 0.18 on a 1–10 scale. Although it is small, it needs to be considered since 80% of the data points were 7 or higher. Female students rate courses taught solely by female teachers slightly higher than courses where only male teachers are involved, but the size of the effect is negligible. As before, the largest effects on the overall rating of courses are found in the number of students in the course (an effect size of 0.40), and whether the student passed the course (an effect size of 0.6).

When looking at the results from the modelling of the teachers' part and the course part (6 models in total), a similar pattern appears in the overall rating of teachers and courses; that is, the gender combination resulting in the lowest score, on an average, in all the cases is male students grading female teachers or courses taught solely by female teachers. As mentioned previously, the effect size is small.

To investigate what proportion of the total variation in the SET scores can be accounted for by the fixed effects and the random effects, the sizes of the random effect variances (student, course, and teacher) were observed. A minute amount of the total variability can be attributed to the fixed effects, 2.5% for the teacher ratings and 3.5% for the course ratings. This indicates that the effect of gender, age, and other fixed effects is minimal. The variability between students is low as well, explaining 14.8% and 17.2% variability in the ratings. A closer look at the course ratings, 23.3% of the variability is due to the difference between courses. This number is smaller for the teacher rating, 10.1%; however, in that case, 13.3% of the total variability is due to the difference between teachers. The largest variability in both cases is unexplained, 57–59% of the total variability. Similar results were found for the teachers' part and the course part.

Analysis of the SET results – open questions

A total of 33 different types of personal comments were identified. At least 10 of those were frequent enough to allow for statistical analysis. The logistic model described above was applied to all of them as well as the separate models for the supportive and the critical comments for each type of comment. The most frequent personal comments were about performance (supportive), how entertaining the teaching was (supportive),

and how well the teacher explained the teaching material (supportive). A summary of frequencies is presented in Table 3.

As outlined above, the comments analysed are on the teacher and not on the teaching methods or student learning. These can further be divided into two broad categories, comments on how good or bad the teacher is on putting on a show 'explaining, performing, entertaining and knowledge', and those related to personality: 'response, demeanour, how organized the teacher is, their approach towards the student learning, use of voice, and support and connection'. Comments in the first category or the 'show' constituted the major proportion of the comments.

The frequencies are shown here as ratios to account for the difference in the number of comments made by male and female students. In the first category, female students seem to put more emphasis on the *explanations* the teachers provided while the male students put equal focus on *explanations* and *performance*. For both types of comments, the majority of the comments are positive. Both male and female students place the same emphasis on *entertainment*, and both genders are much more likely to mention *entertainment* positively. When it comes to *knowledge,* male students place a greater emphasis on the knowledgeability of the teacher, both as a criticism as well as positively. For the first category, the comments are relatively more positive than negative.

For the second category, where the comments are focused more on the teacher's personality the comments are, however, mostly negative. Male students place more focus on criticism of how *organized* the teacher is than female students who are slightly more positive on this aspect. Male students are more critical of the *approach* than female students. Female students are, however, much more critical when it comes to use of voice than male students. Although the portion of comments on support and connections are only a small share of the overall comments, it is evident that women are more likely to comment on support and connection both positively and critically (Table 4).

In the category of showmanship, it is interesting to see that while female students were more likely to write positive comments on *explanation*, male teachers were likely to receive more positive than critical comments on *explanation*. This difference is statistically significant (5%). Female teachers, in turn, were more likely to receive positive comments on *performance* than male teachers, and this difference is statistically significant (10%). There is almost no difference between male and female teachers for the aspect of entertainment. But the patterns for knowledge are the same as for *explanation*, as

Table 3. The frequencies of the content of personal comments, based on the gender of the student.

	Supportive		Critical		
By student gender	Male	Female	Male	Female	Total
Explanations	23%	25%	14%	16%	22%
Performance	23%	22%	13%	11%	20%
Entertaining	19%	20%	4%	5%	16%
Knowledge	14%	10%	9%	7%	10%
Response*	4%	6%	14%	14%	7%
Demeanor	6%	5%	12%	13%	7%
Organization	4%	5%	18%	15%	7%
Approach*	4%	3%	8%	4%	4%
Use of voice	1%	1%	8%	14%	3%
Support and connection	2%	3%	0%	1%	2%

*P–value<0.05.

Table 4. The frequencies of the content of personal comments, based on the gender of the teacher.

	Supportive		Critical		
By teacher gender	Male	Female	Male	Female	Total
Explanations**	25%	23%	17%	13%	22%
Performance*	21%	24%	12%	12%	20%
Entertaining	20%	19%	4%	5%	16%
Knowledge	12%	10%	8%	7%	10%
Response	6%	5%	12%	15%	7%
Demeanor	6%	6%	12%	15%	7%
Organization**	4%	5%	16%	16%	7%
Approach	3%	3%	4%	6%	4%
Use of voice	1%	1%	14%	11%	3%
Support and connection	1%	4%	0%	1%	2%

*P-value<0.05; **P-value<0.01.

male teachers receive more comments on their knowledge of the subject both supportive and critical.

When it comes to the comments that are related to the teacher's personality, the difference based on the gender of the teacher is small, but female teachers seem to get slightly more critical comments. Female teachers do get statistically significantly (5%) more positive comments on how *organized* they are. The only aspect where male teachers get more critical comments than females is *use of voice*, and a majority of these are from female students. Although very few comments are made on *support and connection*, it is interesting that most of them are positive and directed at female teachers. It stands out that no male teacher received a critical comment on *support and connection*.

Discussion

In this article, we asked if gender differences exist in students' answers to SET in a country that is regarded as a frontrunner in gender equality in the world (World Economic Forum, 2021). Thereby we participate in a discussion on whether SET reflects and possibly recreates gender discrimination among university teachers. The analyses show that the gender difference is small, however, male students tend to rate female teachers lower than male teachers, and the difference holds for both measures related to the teacher and the organization of the course. Female students, on the other hand, give courses similar grades as male students give courses solely taught by male teachers. Even though the effects are small, there is a bias present, since the same pattern is observed when looking at the results from the modelling of the teachers' part and the course part (six models in total). That is, on average, the gender combination resulted in the lowest score. In all the cases male students graded female teachers or courses taught solely by female teachers lower.

An analysis of the answers to the open questions in SET, the differences in the types of comments on the teacher based on the gender of the student as well as the gender of teacher further show differences. This indicates that the student's expectations of male and female teachers differ. Male teachers are perceived as knowledgeable and female teachers as caring. Interestingly, the open questions show that male teachers get more critical comments than their female counterparts, particularly from male students, who nevertheless still rate the male teachers higher than a female teacher in the overall score.

In light of this, we cannot conclude that there is no gender bias in SET at the University of Iceland. The difference is small but repeatedly significant. These results are in line with the findings of Boring and Ottoboni (2016) and Pounder (2007), who found indications of gender discrimination in SET. It is certainly discouraging that this gender difference persists in a country like Iceland that scores highly in the Gender Gap Index (World Economic Forum, n.d.) and where gender equality has been used as a brand in the 'politics of reputation' (Einarsdóttir, 2020, p. 148). If gender bias exists among people who are likely to be well acquainted with the discussion on gender bias and stereotypes, as are university students and academic staff in Iceland, gender bias in SET, in general, is hard to dismiss.

Understanding gender bias in SET is important, due to the role it plays in promotion and tenure decisions (Algozzine et al., 2004). In this way, it might contribute to the leaky pipe phenomenon (Anders, 2004; Gasser & Shaffer, 2014) and the gender disparity in the top-level academic positions (Carlsson et al., 2020; Cech & Blair-Loy, 2019). However, one limitation of the study is that we don't know whether the gender difference found in our data is due to a difference in teaching styles between the genders (Centra & Gaubatz, 2000; Punyanunt-Carter & Carter, 2015) as such analysis was not a part of our study. It is also worth mentioning that there is no comparative analysis for the SET used at UI, eliminating direct comparison of the bias observed in SET at the university to other countries. Despite this, an important strength of this research is the use of both quantitative and qualitative data to analyse SET. Another strength is that this is, to our best knowledge, the first study to be conducted on the whole population at a particular university over five years.

Given that a lack of gender diversity in higher academic positions is still a fact and that female teachers' career promotion 'pipes' seem to be 'leakier' than those of their male counterparts (Anders, 2004; Gasser & Shaffer, 2014), there is an urgent need to gain a better understanding of the reasons. This research, conducted in a country that is comparatively better in terms of general gender equality (World Economic Forum, 2021), shows that there are still traces of gendered working culture, where female and male teachers are addressed differently. We cannot, based on this research, conclude whether these differences affect career promotion among women and men at the university. However, we argue that this may well be the case. Therefore, we encourage university authorities around the world, in countries where gender equality is not rated as highly as in the country studied here, to take these findings seriously when aiming to improve gender equality in higher education.

Conclusions

Based on this research there is a gender bias in SET in Iceland, despite it being the frontrunner in gender equality. The takeaway is that for the male teachers to be addressed as good teachers it is enough to explain the subject matter well and talk in a clear voice, while female teachers are rated by their care and service to students. This could potentially have a notable effect on how female and male teachers develop in their academic work and the (re)creation of gendered expectations or stereotypes (Barrow et al., 2020; Lipton, 2021).

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Margret Sigrun Sigudardottir http://orcid.org/0000-0001-8719-0430
Gudbjorg Linda Rafnsdottir http://orcid.org/0000-0003-2662-5773
Anna Helga Jónsdóttir http://orcid.org/0000-0002-0614-0309
Dadi Mar Kristofersson http://orcid.org/0000-0003-3234-6009

References

Algozzine, B., Gretes, J., Flowers, C., Howley, L., Beattie, J., Spooner, F., Mohanty G., & Bray, M. (2004). Student evaluation of college teaching: A practice in search of principles. *College Teaching*, *52*(4), 134–141. https://doi.org/10.3200/CTCH.52.4.134-141

Anders, S. (2004). Why the academic pipeline leaks: Fewer men than women perceive barriers to becoming professors. *Sex Roles*, *51*(9/10), 511–521. https://doi.org/10.1007/s11199-004-5461-9

Barrow, M., Grant, B., & Xu, L. (2020). Academic identities research: Mapping the field's theoretical frameworks. *Higher Education Research & Development*, *42*(2), 240–253. https://doi.org/10.1080/07294360.2020.1849036

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, *145*, 27–41. https://doi.org/10.1016/j.jpubeco.2016.11.006

Boring, A., & Ottoboni, K. (2016). *Student evaluations of teaching (mostly) do not measure teaching effectiveness*. ScienceOpen Research.

Carlsson, M., Finseraas, H., Midtbøen, A. H., & Rafnsdóttir, G. L. (2021). Survey experiment in the Nordic region reveals bias against men in academic evaluations. *European Sociological Review*. https://doi.org/10.1093/esr/jcaa050

Casey, B. H. (2009). The economic contribution of PhDs. *Journal of Higher Education Policy and Management*, *31*(3), 219–227. https://doi.org/10.1080/13600800902974294

Cech, E. A., & Blair-Loy, M. (2019). The changing career trajectories of new parents in STEM. *PNAS*, *116*, 10. https://doi.org/10.1073/pnas.1810862116

Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education*, *71*(1), 17–33. https://doi.org/10.2307/2649280

Clayson, D. E. (2020). Student perception of instructors: The effect of age, gender and political leaning. *Assessment & Evaluation in Higher Education*, *45*(4), 607–616. https://doi.org/10.1080/02602938.2019.1679715

Clayson, D. (2022). The student evaluation of teaching and likability: What the evaluations actually measure. *Assessment & Evaluation in Higher Education*, *47*(2), 313–326. https://doi.org/10.1080/02602938.2021.2011133

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Denson, N., Loveday, T., & Dalton, H. (2010). Student evaluation of courses: What predicts satisfaction? *Higher Education Research & Development*, *29*(4), 339–356. https://doi.org/10.1080/07294360903394466

Einarsdóttir, ÞJ. (2020). All that glitters is not gold: Shrinking and bending gender equality in rankings and nation branding. *NORA - Nordic Journal of Feminist and Gender Research*, *28*(2), 140–152. https://doi.org/10.1080/08038740.2020.1745884

Fan, Y., Shepherd, L. J., Slavich, E., Waters, D., Stone, M., Abel, R., … Johnston, E. L. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PLoS One*, *14*(2), e0209749. https://doi.org/10.1371/journal.pone.0209749

Gasser, C. E., & Shaffer, K. S. (2014). Career development of women in academia: Traversing the leaky pipeline. *The Professional Counselor*, *4*(4), 332–352. https://doi.org/10.15241/ceg.4.4.332

Heffernan, T. (2021). Sexism, racism, prejudice, and bias: A literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment & Evaluation in Higher Education*, 1–11. https://doi.org/10.1080/02602938.2021.1888075

Heijstra, T. M., Steinthorsdóttir, F. S., & Einarsdóttir, T. (2017). Academic career making and the double-edged role of academic housework. *Gender and Education*, *29*(6), 764–780. https://doi.org/10.1080/09540253.2016.1171825

Hertz, N. (2016). Why Iceland is the best place in the world to be a women. https://www.theguardian.com/lifeandstyle/2016/oct/24/iceland-best-place-to-be-women-equal-gender-maternity

Hoorens, V., Dekkers, G., & Deschrijver, E. (2021). Gender bias in student evaluations of teaching: Students' self-affirmation reduces the bias by lowering evaluations of male professors. *Sex Roles*, *84*(1), 34–48. https://doi.org/10.1007/s11199-020-01148-8

Kilpatrick, R. (2017). Iceland has become the first country to officially require gender pay equality. http://fortune.com/2017/03/09/women-iceland-pay-equality-gender/

Kulik, J. A. (2001). Student ratings: Validity, utility, and controversy. *New Directions for Institutional Research*, *2001*(109), 9–25. https://doi.org/10.1002/ir.1

Le Feuvre, N. (2015). Contextualizing Women's Academic Carrers in Cross-National Perspective. GARCIA working paper n. 3, University of Trento (ISBN 978-88-8443-637-5). Retrieved April 23, 2021, from http://garciaproject.eu/wp-content/uploads/2015/10/GARCIA_report_wp3.pdf

Lipton, B. (2021). Academics' dress: Gender and aesthetic labour in the Australian university. *Higher Education Research & Development*, *40*(4), 767–780. https://doi.org/10.1080/07294360.2020.1773767

Mitchell, K. M., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, *51*(3), 648–652. https://doi.org/10.1017/S104909651800001X

Nielsen, M. W. (2017). Scandinavian approaches to gender equality in academia: A comparative study. *Scandinavian Journal of Educational Research*, *61*(3), 295–318. https://doi.org/10.1080/00313831.2016.1147066

Peterson, D. A., Biederman, L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLoS One*, *14*(5), e0216241. https://doi.org/10.1371/journal.pone.0216241

Pinheiro, J. C., & Bates, D. M. (2006). *Mixed-effects models in S and S-PLUS*. Springer.

Pounder, J. S. (2007). Is student evaluation of teaching worthwhile? *Quality Assurance in Education*, *15*(2), 178–191. https://doi.org/10.1108/09684880710748938

Punyanunt-Carter, N., & Carter, S. L. (2015). Students' gender bias in teaching evaluations. *Higher Learning Research Communications*, *5*(3), 28–37. https://doi.org/10.18870/hlrc.v5i3.234

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin*, *26*(11), 1329–1342. https://doi.org/10.1177/0146167200263002

Staub, M., & Heijstra, T. M. (2021). This would never be possible if not for our team play: An analysis of Icelandic doctorate earnings from a gender perspective. *Journal of Family and Economic Issues*. https://doi.org/10.1007/s10834-021-09788-y

Staub, M., & Rafnsdóttir, G. L. (2020). Gender, agency, and time use among doctorate holders: The case of Iceland. *Time and Society*, *29*(1), 143–165. https://doi.org/10.1177/0961463X19884481

Strauss, A., & Corbin, J. M. (1997). *Grounded theory in practice*. Sage.

Teigen, M., & Skjeie, H. (2017). The Nordic gender equality model. In O. Knutsen (Ed.), *The Nordic models in political science. Challenged but still viable?* (pp. 125–147). Fagbokforlaget.

Thébaud, S., & Taylor, C. J. (2021). The specter of motherhood: Culture and the production of gendered career aspirations in science and engineering. *Gender & Society*, *35*(3), 395–421. https://doi.org/10.1177/08912432211006037

The Economist. (n.d.). The glass-ceiling index. https://infographics.economist.com/2017/glass-ceiling/
World Economic Forum. (2021). The Global gender gap report 2021, Geneva. Retrieved July 21, 2021, from https://www.weforum.org/reports/global-gender-gap-report-2021
World Economic Forum. (n.d.). https://www.weforum.org/agenda/2017/11/why-iceland-ranks-first-gender-equality/