The Role of Gender in Students' Ratings of Teaching Quality in Computer Science and Environmental Engineering

Journal Item

The Role of Gender in Students' Ratings of Teaching Quality in Computer Science and Environmental Engineering

Linda Price, Ingrid Svensson, Jonas Borell, and John T. E. Richardson

***Abstract*—Students' ratings of teaching quality on course units in computer science and environmental engineering at a large Swedish university were obtained using the Course Experience Questionnaire; 8,888 sets of ratings were obtained from men and 4,280 sets were obtained from women over ten academic years. There were differences in the ratings given by students taking the two programs; in particular, teachers tended to receive higher ratings in subjects that were less typical for their gender than in subjects that were more typical for their gender. There were differences in the ratings given to male and female teachers, differences in the ratings given by male and female students, and interactions between these two effects. There was no systematic trend for students to give different ratings to teachers of the same gender as themselves compared with teachers of the other gender. Nevertheless, without exception even the statistically significant effects were small in magnitude and unlikely to be of theoretical or practical importance. It is concluded that the causes of differences in the career progression of male and female teachers in engineering education need to be sought elsewhere.**

***Index Terms*—Computer science, engineering students, environmental engineering, gender, teaching evaluations.**

I. INTRODUCTION

Women have historically been under-represented in engineering education, both as students and as teachers. In the UK, for instance, women constituted 56.2% of all undergraduate students in the academic year 2014–2015 but only 14.5% of those taking programs in engineering and technology [1]. This situation appears not to have changed over the last 30 years [2]. Even so, there are variations across different engineering programs: For example, at the university where the present study was carried out, female students constituted just 7% of the enrollment in computer science but around 60% of the enrollment in environmental engineering.

Manuscript received

L. Price is head of Technology Enhanced Learning, Kingston University, Kingston upon Thames, Surrey KT1 1LQ, UK (e-mail: Linda.Price@kingston.ac.uk).

I. Svensson is with the Department of Biomedical Engineering, Lund University, SE-221 00 Lund, Sweden (e-mail: ingrid.svensson@bme.lth.se).

J. Borell is with Ergonomics and Aerosol Technology, Lund University, SE-221 00 Lund, Sweden (e-mail: jonas.borell@design.lth.se).

J. T. E. Richardson is with the Institute of Educational Technology, The Open University, Milton Keynes MK7 6AA, UK (e-mail: John.T.E.Richardson@open.ac.uk).

Digital Object Identifier

In the past, such disparities have been ascribed to genetically determined differences in men's and women's mathematical and spatial abilities. However, the differences in question tend to be small and inconsistent and often result from biases in the sampling of the participants [3]. The substantial under-representation of women in engineering education is more likely to be due to differential socialization of boys and girls [4], [5], leading to different choices of options in high school [6] and different drop-out rates from engineering programs at university [7]. There are also gender differences in the career progression of engineering teachers generally favoring men over women [8], [9].

This study was concerned with student's perceptions of teaching quality on programs in computer science and environmental engineering, as measured by their feedback in questionnaire surveys at the end of each course unit.

A. Gender Differences in Students' Evaluations of Teaching

There is an extensive literature on the role of gender in student feedback, but previous studies have been confined to traditional subjects such as the sciences, the social sciences, and education rather than engineering. In an experimental study, Harris [10] asked students to evaluate a fictitious professor of engineering, but the students themselves were taking a psychology program and are unlikely to have had first-hand knowledge of engineering education.

Feldman reported two reviews of the research literature on how students perceived male and female teachers. The first described the results of laboratory research and other artificial experiments [11]. The majority of studies had found no difference in students' overall evaluations of male and female teachers, although, where differences had been found, male teachers tended to receive more positive ratings than female teachers. In general, male and female teachers were rated in a broadly similar way by male and female students.

Feldman's second review described the results of students' actual evaluations of their classroom teachers [12]. Once again, the majority of studies had found no difference in students' overall evaluations of male and female teachers. In this case, where differences had been found, more of them favored female teachers than favored male teachers. However, the average point-biserial correlation coefficient between the teachers' gender and the students' ratings across all the relevant studies was only 0.02, which Feldman argued was "so small as to be substantively negligible" (p. 177). There was a slight tendency for students to rate teachers of the same gender as themselves more highly than teachers of the other gender, although this varied across different studies.

Feldman [11] noted that in five studies involving artificial experiments students' overall evaluations had been obtained across different academic subjects. In all five studies, there was no significant interaction between the teachers' gender and the academic subject, implying that students' relative evaluations of male and female teachers did not vary across different academic subjects. Feldman [12] found just two studies in which this had been explored in classroom evaluations. In one, the students had tended to give higher ratings to their former teachers in academic subjects that were *typical* for their gender [13]. However, in the second, the students had tended to give higher ratings to their current teachers in subjects that were *atypical* for their gender [14].

Feldman had confined his reviews to research published in the USA and Canada. Subsequently, other relevant literature has appeared from around the world. Nevertheless, the findings broadly confirm Feldman's conclusions with regard to the effect of teachers' gender on students' ratings, the effect of students' gender on students' ratings, and the interaction between them: Each of these effects is typically small, inconsistent in direction, and often nonsignificant. Indeed, two studies even found no significant differences between students' ratings of male and female teachers on the website RateMyProfessors.com [15], [16].

Even so, a recent study has revived interest in these issues. Boring [17] obtained the 22,665 ratings given by male and female students who had taken mandatory first-year course units in the social sciences at a French university over five successive academic years. These students had been assigned in an unsystematic way to seminar groups of about 20 students led by male and female teachers, and at the end of each quarter they rated the teachers who had led their seminar groups.

Boring found that male students tended to rate male teachers more positively than they rated female teachers, whereas female students gave similar ratings to male and female teachers. Male teachers who were rated by male students tended to receive the highest ratings. Boring also noted that all students tended to rate male teachers more highly on their class leadership and being up-to-date with current issues, while female teachers were rated more highly on the more time-consuming activities of course preparation and organization. Boring suggested that this might explain why female teachers seemed to spend more time on teaching and less time on research than male teachers.

Unfortunately, there are a number of problems with this study. First, the questionnaire used to obtain students' ratings consisted of just nine items about specific dimensions, plus an item concerned with overall satisfaction. The instrument was constructed in-house, and it was apparently implemented without evaluating its reliability or validity. Boring's report contains information about the statistical significance of the findings, but it does not present measures of effect size. Indeed, many of the differences that she identified are fairly small in magnitude and may only have achieved statistical significance because of the very large sample size.

B. Context and Aims of the Present Study

An opportunity arose to investigate the ratings given by male and female students on course units taught by male and female teachers in the engineering faculty of a large university in Sweden. For many years, the faculty had routinely obtained feedback from students using the Course Experience Questionnaire (CEQ). This was originally devised by Ramsden [18] as a performance indicator for monitoring the quality of teaching on programs at Australian universities. A version of the CEQ containing 23 items was implemented in a national survey of students who had graduated in 1992 [19], and the exercise was repeated annually thereafter. The psychometric properties of different versions of the CEQ are well established, and it has been used with both graduates and currently enrolled students and on both entire programs and individual course units [20].

The version of the CEQ used in the Australian graduate surveys contained the five scales shown in Table I. The respondents were instructed to indicate their level of agreement or disagreement (along a scale from "definitely agree," scoring 5, to "definitely disagree," scoring 1) with each statement as a description of their program of study. Some of the items referred to positive aspects, but other items referred to negative aspects and were to be coded in reverse (so that "definitely agree" is coded as 1, and "definitely disagree" is coded as 5). The scales contain varying numbers of items, and respondents are therefore assigned scale scores by averaging the coded responses across the relevant items. As a result, the scale scores themselves also range from 1 to 5.

The Australian graduate surveys identified apparent overall differences between male and female students in their ratings of their programs. However, the authors of the reports from these surveys were at pains to point out that these differences might simply reflect the enrollment of men and women on programs in different disciplines with different teaching practices and different assessment requirements. In other words, the differences in their ratings might arise from men and women choosing different programs rather than from their different genders.

TABLE I
SCALES AND EXAMPLE ITEMS FOR THE 23-ITEM CEQ

Scale	Example item
Appropriate Assessment	Staff here seem more interested in testing what we have memorized than what we have understood.*
Appropriate Workload	The sheer volume of work to be got through means you can't comprehend it all thoroughly.*
Clear Goals and Standards	You usually have a clear idea of where you're going and what's expected of you in this course.
Generic Skills	This course has helped me to develop my problem-solving skills.
Good Teaching	Teaching staff here normally give helpful feedback on how you are going.

Items with asterisks are negatively worded and are to be coded in reverse.

In contrast, studies that have compared ratings on the CEQ given by students taking the *same* course units or programs found no significant difference between male and female students in their CEQ scores [21], [22]. Research studies involving large samples of students taking the same course units have found statistically significant gender differences, but the effects are small in magnitude and inconsistent from one study to another [23]. Apart from one which investigated postgraduate students taking computer programming [22], none of these studies involved engineering students.

It would appear that no studies have compared CEQ ratings given by students taking course units led by male teachers and course units led by female teachers, or CEQ ratings given by students taking course units led by teachers of the same gender and course units led by teachers of the other gender. Accordingly, the aim of this study was to investigate the CEQ responses given by male and female students taking course units in engineering led by male and female teachers. The two programs mentioned earlier were chosen for comparison: computer science, which typically attracted small numbers of female students, and environmental engineering, which typically attracted large numbers of female students. Both were of five years' duration and led to a Master's degree:

The study addressed four research questions:

1. Does the gendered nature of the teaching–learning context affect students' ratings of their course units on the CEQ?
2. Do male and female *teachers* receive different ratings from their students?
3. Do male and female *students* give different ratings to their teachers?
4. Do students give different ratings to teachers of the *same* gender as themselves compared with teachers of the *other* gender?

II. Method

A. Procedure

The academic year in the relevant faculty was divided into two semesters, and each semester was divided into two study periods or quarters. The course units making up each program were all taught for a single quarter. The CEQ was administered at the end of each presentation, initially on paper but more recently online. Respondents were asked to declare their age and gender, but otherwise the surveys were anonymous. The response rate was typically about 45%.

B. Data Analysis

A multivariate analysis of variance was carried out on the students' scores on the five scales of the CEQ. Each student may have contributed responses on several course units; however, since their responses were anonymous, each set of scores had to be treated as an independent observation. The analysis employed the independent variables of the students' study program, the teachers' gender, the students' gender, the students' course unit, and the quarter in which the unit was taught. The gender of the teacher with overall responsibility for the course unit was used if more than one teacher was involved in its presentation. The course units were nested within the programs of study. All possible interaction terms were computed among the students' program, the teachers' gender, and the students' gender. However, since most course units were not taught in every quarter, it was not feasible to calculate any further interactions.

For each effect and interaction, Wilks' lambda (Λ) is reported as the multivariate statistic in an omnibus test with its associated *F* test, followed by univariate tests on each of the CEQ scales. Interactions that were statistically significant were investigated further by means of tests on simple main effects. For both multivariate and univariate tests, partial eta squared (η^2) is reported as a measure of effect size. This measures the proportion of the total variance in a dependent variable that is associated with the membership of different groups defined by an independent variable or interaction when the effects of other independent variables and interactions have been partialed out. Cohen [24, p. 280] proposed that values of partial η^2 of 0.0099, 0.0588, and 0.1379 would reflect "small," "medium," and "large" effects, respectively.

III. Results

The data covered 455 presentations of 75 course units over 39 quarters between 2004–2005 and 2013–2014. Over the 10 academic years, enrollment on the computer science program varied between 408 and 602 students, with a mean of 491.4 students, of whom 93.1% were men and 6.9% were women. Enrollment on the environmental engineering program varied between 223 and 295 students, with a mean of 260.5 students, of whom 39.8% were men and 60.2% were women.

Of the 455 presentations, 77.1% were taught by men (81.3% in computer science and 72.2% in environmental engineering.) There were 13,168 complete sets of responses to the CEQ, along with information about the respondents' gender; 7,588 sets were provided by students taking course units in computer science and 5,580 were provided by students taking course units in environmental engineering; 8,888 sets were provided by men and 4,280 were provided by women; 10,153 sets were provided by students taught by men and 3,015 were provided by students taught by women.

A. Study Program by Teachers' Gender

There was a small but highly significant multivariate effect of the students' study program, Wilks' $\Lambda = 0.990$, $F(5, 13041) = 25.94$, $p < 0.001$, partial $\eta^2 = 0.010$. There was also a small but highly significant multivariate effect of the teachers' gender, Wilks' $\Lambda = 0.997$, $F(5, 13041) = 8.04$, $p < 0.001$, partial $\eta^2 = 0.003$. However, these effects were qualified by a small but highly significant multivariate interaction between the effects of the students' study program and the teachers' gender, Wilks' $\Lambda = 0.998$, $F(5, 13041) = 5.69$, $p < 0.001$, partial $\eta^2 = 0.004$. The mean ratings given to the male and female teachers by the computer science students and the environmental engineering students (and their standard errors) are shown in Table II, adjusted for the effects of the other variables in the research design.

TABLE II
CEQ SCALE SCORES BY PROGRAM AND TEACHERS' GENDER

Scale	Male teachers		Female teachers		
	M	*SE*	*M*	*SE*	*p*
Computer science					
Appropriate Assessment	3.72	0.03	3.79	0.05	0.18
Appropriate Workload	2.99	0.03	3.15	0.05	0.00
Clear Goals and Standards	3.20	0.03	3.52	0.05	0.00
Generic Skills	3.04	0.03	3.09	0.04	0.29
Good Teaching	3.15	0.03	3.17	0.05	0.62
Environmental engineering					
Appropriate Assessment	3.75	0.03	3.78	0.04	0.59
Appropriate Workload	3.04	0.03	2.85	0.05	0.00
Clear Goals and Standards	3.27	0.03	3.13	0.04	0.01
Generic Skills	3.30	0.02	3.24	0.04	0.17
Good Teaching	3.42	0.03	3.19	0.04	0.00

Values of *p* are rounded to two decimal places and indicate the significance of the differences between the mean scores of the male and female teachers.

TABLE III
CEQ SCALE SCORES BY PROGRAM AND STUDENTS' GENDER

Scale	Male students		Female students		
	M	*SE*	*M*	*SE*	*p*
Computer science					
Appropriate Assessment	3.77	0.03	3.74	0.05	0.51
Appropriate Workload	3.17	0.03	2.98	0.05	0.00
Clear Goals and Standards	3.40	0.03	3.32	0.05	0.05
Generic Skills	3.06	0.02	3.06	0.04	0.97
Good Teaching	3.20	0.02	3.12	0.04	0.05
Environmental engineering					
Appropriate Assessment	3.73	0.03	3.80	0.03	0.10
Appropriate Workload	3.02	0.03	2.87	0.03	0.00
Clear Goals and Standards	3.24	0.03	3.16	0.03	0.00
Generic Skills	3.28	0.03	3.27	0.03	0.54
Good Teaching	3.34	0.03	3.28	0.03	0.01

Values of *p* are rounded to two decimal places and indicate the significance of the differences between the mean scores of the male and female students.

There were significant univariate interactions on the Appropriate Workload scale, $F(1, 13045) = 24.20$, $p < 0.001$, partial $\eta^2 = 0.002$, the Clear Goals scale, $F(1, 13045) = 42.00$, $p < 0.001$, partial $\eta^2 = 0.003$, and the Good Teaching scale, $F(1, 13045) = 14.72$, $p < 0.001$, partial $\eta^2 = 0.001$. No other univariate interactions were statistically significant.

Post hoc tests using simple main effects showed that the computer science students gave higher ratings to female teachers than to male teachers on the Appropriate Workload scale, $F(1, 13045) = 9.61$, $p = 0.002$, partial $\eta^2 = 0.001$, and the Clear Goals scale, $F(1, 13045) = 36.67$, $p < 0.001$, partial $\eta^2 = 0.003$. However, the environmental engineering students gave higher ratings to male teachers than to female teachers on the Appropriate Workload scale, $F(1, 13045) = 14.60$, $p < 0.001$, partial $\eta^2 = 0.001$, the Clear Goals scale, $F(1, 13045) = 8.61$, $p = 0.003$, partial $\eta^2 = 0.001$, and the Good Teaching scale, $F(1, 13045) = 25.30$, $p < 0.001$, partial $\eta^2 = 0.002$.

Further post hoc tests showed that male teachers were given higher ratings by environmental engineering students than by computer science students on the Appropriate Assessment scale, $F(1, 13045) = 5.18$, $p = 0.023$, partial $\eta^2 = 0.000$, the Appropriate Workload scale, $F(1, 13045) = 159.03$, $p < 0.001$, partial $\eta^2 = 0.012$, the Generic Skills scale, $F(1, 13045) = 61.36$, $p < 0.001$, partial $\eta^2 = 0.005$, and the Good Teaching scale, $F(1, 13045) = 8.52$, $p = 0.004$, partial $\eta^2 = 0.001$. Female teachers were also given higher ratings by environmental engineering students than by computer science students on the Generic Skills scale, $F(1, 13045) = 31.47$, $p < 0.001$, partial $\eta^2 = 0.002$. However, they were given higher ratings by computer science students than by environmental engineering students on the Appropriate Workload Scale, $F(1, 13045) = 53.34$, $p < 0.001$, partial $\eta^2 = 0.004$, and the Clear Goals scale, $F(1, 13045) = 23.80$, $p < 0.001$, partial $\eta^2 = 0.002$. No other post hoc tests were statistically significant.

B. Study Program by Students' Gender

The mean ratings given by the male and female students taking course units in computer science and environmental engineering (and their standard errors) are shown in Table III, adjusted for the effects of the other variables in the research design. There was a small but highly significant multivariate effect of the students' gender, Wilks' $\Lambda = 0.996$, $F(5, 13041) = 11.04$, $p < 0.001$, partial $\eta^2 = 0.004$. Univariate tests showed that the male students produced higher ratings than the female students on the Appropriate Workload scale, $F(1, 13045) = 45.80$, $p < 0.001$, partial $\eta^2 = 0.003$, the Clear Goals scale, $F(1, 13045) = 10.96$, $p = 0.001$, partial $\eta^2 = 0.001$, and the Good Teaching scale, $F(1, 13045) = 9.06$, $p = 0.003$, partial $\eta^2 = 0.001$. No other univariate effects were statistically significant.

The multivariate interaction between the effects of the students' study program and their gender was not statistically significant, Wilks' $\Lambda = 1.000$, $F(5, 13041) = 0.93$, $p = 0.460$, partial $\eta^2 = 0.000$, and none of the univariate interactions was significant. These results imply (and the results of post hoc tests confirm) that the pattern of results shown in Table III was broadly similar in computer science students and environmental engineering students.

C. Study Program by Teachers' Gender by Students' Gender

The mean ratings given to male and female teachers by male and female students taking course units in computer science and environmental engineering (and their standard errors) are shown in Tables IV and V, adjusted for the effects of the other variables in the research design. There was a small but highly significant multivariate interaction between the effects of the teachers' gender and the students' gender, Wilks' $\Lambda = 0.998$, $F(5, 13041) = 5.69$, $p < 0.001$, partial $\eta^2 = 0.002$. There were significant univariate interactions on the Appropriate Workload scale, $F(1, 13045) = 9.13$, $p = 0.003$, partial $\eta^2 = 0.001$, the Clear Goals scale, $F(1, 13045) = 19.83$, $p < 0.001$, partial $\eta^2 = 0.002$, and the Good Teaching scale, $F(1, 13045) = 19.14$, $p < 0.001$, partial $\eta^2 = 0.001$. No other univariate interactions were significant.

Post hoc tests using simple main effects showed that the male students gave higher ratings to male teachers than to female teachers on the Appropriate Workload scale, $F(1, 13045) = 5.77$, $p = 0.016$, partial $\eta^2 = 0.000$, and on the Good Teaching scale, $F(1, 13045) = 36.27$, $p < 0.001$, partial $\eta^2 = 0.003$. However, the female students gave higher ratings to female teachers than to male teachers on the Clear Goals scale,

TABLE IV
CEQ SCALE SCORES BY TEACHERS' GENDER AND STUDENTS' GENDER: COMPUTER SCIENCE

Scale	Male students		Female students		
	M	*SE*	*M*	*SE*	*p*
Male teachers					
Appropriate Assessment	3.73	0.02	3.71	0.04	0.66
Appropriate Workload	3.14	0.02	2.84	0.04	0.00
Clear Goals and Standards	3.31	0.02	3.10	0.04	0.00
Generic Skills	3.06	0.02	3.02	0.04	0.25
Good Teaching	3.23	0.02	3.06	0.04	0.00
Female teachers					
Appropriate Assessment	3.81	0.04	3.77	0.08	0.61
Appropriate Workload	3.19	0.04	3.11	0.08	0.28
Clear Goals and Standards	3.49	0.04	3.54	0.08	0.49
Generic Skills	3.07	0.03	3.11	0.07	0.52
Good Teaching	3.16	0.04	3.18	0.07	0.78

Values of *p* are rounded to two decimal places and indicate the significance of the differences between the mean scores of the male and female students.

TABLE V
CEQ SCALE SCORES BY TEACHERS' GENDER AND STUDENTS' GENDER: ENVIRONMENTAL ENGINEERING

Scale	Male students		Female students		
	M	*SE*	*M*	*SE*	*p*
Male teachers					
Appropriate Assessment	3.74	0.03	3.77	0.03	0.21
Appropriate Workload	3.13	0.03	2.94	0.03	0.00
Clear Goals and Standards	3.35	0.03	3.18	0.03	0.00
Generic Skills	3.32	0.03	3.28	0.03	0.05
Good Teaching	3.51	0.03	3.34	0.03	0.00
Female teachers					
Appropriate Assessment	3.73	0.05	3.83	0.05	0.02
Appropriate Workload	2.90	0.05	2.79	0.05	0.02
Clear Goals and Standards	3.12	0.05	3.13	0.05	0.92
Generic Skills	3.23	0.05	3.25	0.04	0.65
Good Teaching	3.17	0.05	3.21	0.04	0.30

Values of *p* are rounded to two decimal places and indicate the significance of the differences between the mean scores of the male and female students.

$F(1, 13045) = 15.29$, $p < 0.001$, partial $\eta^2 = 0.001$.

Further post hoc tests showed that the male teachers were given higher ratings by male students than by female students on the Appropriate Workload scale, $F(1, 13045) = 103.83$, $p < 0.001$, partial $\eta^2 = 0.008$, the Clear Goals scale, $F(1, 13045) = 65.27$, $p < 0.001$, partial $\eta^2 = 0.005$, the Generic Skills scale, $F(1, 13045) = 4.25$, $p = 0.039$, partial $\eta^2 = 0.000$, and the Good Teaching scale, $F(1, 13045) = 59.06$, $p < 0.001$, partial $\eta^2 = 0.005$. However, the female teachers were given higher ratings by male students than by female students only on the Appropriate Workload scale, $F(1, 13045) = 4.57$, $p = 0.032$, partial $\eta^2 = 0.000$. No other tests were statistically significant.

The three-way multivariate interaction between the effects of the students' study program, the teachers' gender, and the students' gender was not statistically significant, Wilks' $\Lambda = 1.000$, $F(5, 13041) = 0.95$, $p = 0.447$, partial $\eta^2 = 0.000$, and none of the univariate interactions was significant. Although there are one or two discrepancies, these results imply (and the results of post hoc tests generally confirm) that the pattern of results shown in Tables IV and V was broadly similar in computer science students and environmental engineering students.

D. Course Unit

There was a large and highly significant multivariate effect of the students' course unit, nested within the two programs, Wilks' $\Lambda = 0.338$, $F(385, 65100) = 41.00$, $p < 0.001$, partial $\eta^2 = 0.195$. There were large and highly significant univariate effects on each of the five scales, $F(77, 13045) \geq 34.57$, $p < 0.001$, partial $\eta^2 \geq 0.169$ in each case.

E. Quarter

There was a small but highly significant multivariate effect of the quarter in which the students' ratings were collected, Wilks' $\Lambda = 0.949$, $F(190, 64760) = 3.59$, $p < 0.001$, partial $\eta^2 = 0.010$. There were small but highly significant univariate effects on each of the five scales, $F(38, 13045) \geq 1.70$, $p \leq 0.005$, partial $\eta^2 \geq 0.005$ in each case. To determine whether these represented systematic chronological variations, the linear and quadratic trends were calculated across the 39 quarters. The linear trend was significant (and positive) for the Generic Skills scale, $F(1, 13045) = 61.19$, $p < 0.001$, partial $\eta^2 = 0.005$, reflecting a small increase between 2004–2005 and 2013–2014, but not on any of the other scales. The quadratic trend was not statistically significant on any of the scales.

IV. Discussion

The CEQ was originally devised to differentiate between the experiences of students who had taken programs in the same subjects across different Australian universities [18]. It has also been used to differentiate between the experiences of students taking different course units [20]. The present study found that there was a highly significant variation across the 75 course units constituting the two degree programs on each of the CEQ's scales. Statistically, these were large effects, accounting for more than one-sixth of the variance in each of the scale scores when the effects of other variables and interactions had been statistically controlled. This implies that they were of both theoretical and practical importance [24].

To address Research Question 1, the pattern of ratings given by students of computer science (which attracted small numbers of female students) differed from those given by students of environmental engineering (where female students constituted the majority). In computer science, female teachers were rated more highly than male teachers on Appropriate Workload and Clear Goals and Standards. In environmental engineering, male teachers were rated more highly than female teachers on Appropriate Workload, Clear Goals and Standards, and Good Teaching. In other words, students tended to give higher ratings to teachers in subjects that were less typical for their gender than to teachers in subjects that were more typical for their gender.

To address Research Question 2, male and female teachers received different ratings from their students on two of the CEQ's scales. Female teachers were rated more highly than male teachers on Clear Goals and Standards, but only by female students, not by male students. Conversely, male teachers were rated more highly than female teachers on Good Teaching, but only by male students, not by female students.

To address Research Question 3, male and female students gave different ratings to their teachers on three of the CEQ's scales. Male students produced higher ratings than female students on Clear Goals and Standards and on Good Teaching, but only for male teachers, not for female teachers. Only in the case of Appropriate Workload did both male and female students produce higher ratings of male teachers than of female teachers.

There were statistically significant interactions between the effects of teachers' gender and of students' gender on three of the CEQ's scales: Appropriate Workload, Clear Goals and Standards, and Good Teaching. However, in none of these cases did both male and female students give different ratings to teachers of the *same* gender compared with teachers of the *other* gender, which answers Research Question 4.

Nevertheless, it must be emphasized that the differences between male and female teachers and between male and female students in this study constituted only small effects that had achieved statistical significance simply by virtue of the large sample size (N = 13,168). Without exception, they accounted for less than 1/200th of the variance in each of the scale scores when the effects of other variables and interactions had been statistically controlled. This implies that they were of little theoretical or practical importance [24].

V. LIMITATIONS

One limitation of this study is that it was concerned with students' ratings of teaching quality in two specific programs in one faculty of engineering. Including programs from more than one institution would have been logistically difficult if not impossible. Nevertheless, it would be valuable if colleagues elsewhere were to report analogous data from their own institutions for comparative purposes. Even so, the inclusion in the present study of two programs with very different profiles in terms of student gender means that one can have some confidence in the generalizability of the results.

The other limitation of this study is that each set of scores had to be treated as an independent observation because the students' responses were anonymous. In theory, it would be of interest to compare the ratings produced by the same students on different course units. In practice, however, there are major ethical issues in requiring students to identify themselves when providing feedback in questionnaire surveys, and this would almost certainly have produced a lower response rate [25]. Even broader ethical issues would be raised if an institution sought to track an individual student's responses to online surveys in the pursuit of "learning analytics" [26].

VI. CONCLUSIONS

To return to the four original research questions, this study has found that the gendered nature of the teaching–learning context affects students' ratings of their course units. Specifically, students tended to give higher ratings on the CEQ to teachers in subjects that were less typical for their gender (women teaching computer science or men teaching environmental engineering) than to teachers in subjects that were more typical for their gender (men teaching computer science or women teaching environmental engineering). This is consistent with the findings of a previous study in the research literature [14].

This study also found that male and female *teachers* received different ratings on the CEQ from their students and that male and female *students* gave different ratings on the CEQ to their teachers. Although there were statistically significant interactions between these two effects, they did not represent a situation in which students gave different ratings on the CEQ to teachers of the *same* gender compared with teachers of the *other* gender. As mentioned, all of these were small effects of little theoretical or practical importance. This is consistent with the conclusions of Feldman's reviews based on research in the USA and Canada [11], [12], as well as with subsequent literature from countries around the world.

In some institutions of higher education and in some systems of higher education, student feedback is nowadays used in the appointment, tenure, promotion, and reward of individual teachers. Boring [17] argued that differences in the ratings given to male and female teachers by male and female students would explain differences in the career progression of male and female teachers. The present findings render such arguments implausible. In environmental engineering, male teachers did tend to receive higher ratings than female teachers, but the effects were small in magnitude and not consistent across the five scales of the CEQ. In computer science, female teachers actually tended to receive higher ratings than male teachers, although once again the effects were small in magnitude and not consistent across the five scales of the CEQ. On the contrary, the causes of differences in the career progression of male and female teachers, at least in engineering education, need to be sought elsewhere.

ACKNOWLEDGMENT

The authors thank Karim Andersson for technical support and for providing access to the data described in this paper. They also thank Beatrice Tucker for helpful comments on the ethics of student feedback surveys.

REFERENCES

[1] Higher Education Statistics Agency, *Statistical First Release 224*. Cheltenham, UK: Higher Education Statistics Agency, 2015.
[2] E. Smith, "Women into science and engineering? Gendered participation in higher education STEM subjects," *Brit. Educ. Res. J., vol.* 37, pp. 993–1014, 2011.
[3] P. J. Caplan, M. Crawford, J. S. Hyde, and J. T. E. Richardson, *Gender Differences in Human Cognition*. New York: Oxford University Press, 1997.
[4] J. R. Campbell and J. S. Beaudry, "Gender gap linked to differential socialization for high-achieving senior mathematics students," *J. Educ. Res*., vol. 91, pp. 140–147, 1998.
[5] J. S. Eccles, J. E. Jacobs, and R. D. Harold, "Gender role stereotypes, expectancy effects, and parents' socialization of gender differences," *J. Soc. Issues*, vol. 46, pp. 183–201, 1990.
[6] K. A. Updegraff, J. S. Eccles, B. L. Barber, and K. M. O'Brien, "Course enrolment as self-regulatory behaviour: Who takes optional high school math courses?" *Learn. Individ. Differ*., vol. 8, pp. 239–259, 1996.
[7] M. A. Beasley and M. J. Fischer, "Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors," *Soc. Psychol. Educ*., vol. 15, pp. 427–

448, 2012.
[8] National Research Council, *Gender Differences at Critical Transitions in the Careers of Science, Engineering, and Mathematics Faculty*. Washington, DC: National Academies Press, 2010.
[9] D. Bilimoria and X. Liang, *Gender Equity in Science and Engineering: Advancing Change in Higher Education*. New York: Routledge, 2012.
[10] M. B. Harris, "Sex role stereotypes and teacher evaluations," *J. Educ. Psychol*., vol. 67, pp. 751–756, 1975.
[11] K. A. Feldman, "College students' views of male and female college teachers: Part I. Evidence from the social laboratory and experiments," *Res. High. Educ*., vol. 33, pp. 317–375, 1992.
[12] _____, "College students' views of male and female college teachers: Part II. Evidence from students' evaluations of their classroom teachers," *Res. High. Educ*., vol. 34, pp. 151–211, 1993.
[13] M. A. Ferber and J. A. Huber, "Sex of student and instructor: A study of student bias," *Amer. J. Sociol*,. vol. 80, 949–963, 1975.
[14] C. R. Tieman and B. Rankin-Ullock, "Student evaluations of teachers: An examination of the effect of sex and field of study," *Teach. Sociol*., vol. 12, 177–191, 1985.
[15] M. A. Mangan and B. Fleck, "Online student evaluation of teaching: Will professor 'hot and easy' win the day?" *J. Excellence Coll. Teach*., vol. 22, pp. 59–84, 2011.
[16] K. C. Theyson, "Hot or not: The role of instructor quality and gender on the formation of positive illusions among students using RateMyProfessors.com," *Pract. Assess. Res. Eval*., vol. 20, no. 4, 2015. Available: http://pareonline.net/getvn.asp?v=20&n=4, Accessed on: Apr 16, 2017
[17] A. Boring, Gender Biases in Student Evaluations of Teachers (Working Paper 2015-13). Paris: Sciences Po, Observatoire Français des Conjonctures Economiques, 2015. Available: http://www.ofce.sciences-po.fr/pdf/dtravail/WP2015-13.pdf, Accessed on: Apr 16, 2017
[18] P. Ramsden, "A performance indicator of teaching quality in higher education: The Course Experience Questionnaire," *Stud. High. Educ*., vol. 16, pp. 129–150, 1991.
[19] J. Ainley and M. Long, *The Course Experience Survey 1992 Graduates*. Canberra: Australian Government Publishing Service, 1994.
[20] J. T. E. Richardson, "What can students' perceptions of academic quality tell us? Research using the Course Experience Questionnaire," in *The Routledge International Handbook of Higher Education*, M. Tight, K. H. Mok, J. Huisman, and C. C. Morphew, Eds. New York: Routledge, 2009, pp. 199–210.
[21] Y. M. Chan and C. M. S. Chan, "Approaches to learning of sub-degree students in Hong Kong," *Res. Educ*., No. 84, 65–78, 2010.
[22] J. T. E. Richardson and L. Price, "Approaches to studying and perceptions of academic quality in electronically delivered courses," *Brit. J. Educ. Technol*., vol. 34, pp. 45–56, 2003.
[23] J. T. E. Richardson, "Investigating the relationship between variations in students' perceptions of their academic environment and variations in study behaviour in distance education," *Brit. J. Educ. Psychol*., vol. 76, pp. 867–893, 2006.
[24] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press, 1969.
[25] M. Shah, C. D. Nair, and J. T. E. Richardson, *Measuring and Enhancing the Student Experience*. Cambridge, MA: Elsevier, 2016.
[26] S. Slade and P. Prinsloo, "Learning analytics: Ethical issues and dilemmas," *Am. Behav. Sci*., vol. 57, pp. 1510–1529, 2013.

Linda Price received the M.Sc. degree in computer science education from the University of Ulster, UK, and the Ph.D. degree in educational technology from The Open University, UK.

She is currently professor of educational technology and head of technology enhanced learning at Kingston University, UK, and professor of engineer education at Lund University, Sweden. Her research interests concern pedagogically driven uses of educational technology in higher education.

Ingrid Svensson received the M.Sc. degree in engineering physics and the Ph.D. degree in solid mechanics from Lund University, Sweden.

She is currently senior lecturer in biomedical engineering and a member of the academic development unit at Lund University, Sweden. Her research interests include all forms of biomechanics.

Jonas Borell received the M.Sc. degree in psychology and the Ph.D. degree in ergonomics from Lund University, Sweden.

He is currently assistant lecturer in ergonomics and aerosol technology at Lund University, Sweden. His research interests include safety management, learning for safety, and human factors in various work environment settings.

John T. E. Richardson received the B.A. degree in philosophy and psychology from the University of Oxford, UK, and the D.Phil. degree in experimental psychology from the University of Sussex, UK.

He is currently professor in student learning and assessment at The Open University, UK. His research interests include various aspects of student learning in higher education.