Study of Impact of Gender on Engagement and Performance of Engineering Students

M. E. Sousa-Vieira, J. C. López-Ardao and M. Fernández-Veiga
atlanTTic. Universidade de Vigo, 36310 Vigo, Spain

Keywords: Gender, Engineering and Technology, Learning Analytics.

Abstract: The gender gap in STEM disciplines manifests itself in several axes. One of these, not covered in depth in the literature, is the likelihood of different academic achievements by men and women in some technology-oriented courses. We address this question in this work, providing a cross-sectional study conducted on an entry-level course on computer networking at the college level. Our findings suggest that, while we do not observe statistically significant difference in the final grades, women perform slightly better in some specific individual tasks and tend to have more intense participation in social learning activities. Interestingly, our results do not confirm the hypothesis that men tend to show higher variance in their achievements than women.

1 INTRODUCTION

The existence of a strong gender gap in STEM (Science, Technology, Engineering and Mathematics) disciplines has long been reported and recognized, and is transversal to many different countries and educational levels. The causes have been thoroughly analyzed from different perspectives, and include several factors: prevalence of stereotypes in education and society, lack of role models, unconscious bias, and disparate perceptions on the work-life balance-imbalance.

The underrepresentation of women in STEM is not only an issue of disparity in (self-taught) preferences or confidence, but it also entails a loss of innovation and competitiveness. There exists solid evidence that groups with more gender diversity are better in problem-solving skills and companies with more balanced work teams are more efficient and productive than those without it (Hundschell et al., 2021). Difference have also been detected between men and women enrolled in STEM undergraduate and graduate degrees (Alexandru et al., 2022). For instance, women who drop from STEM majors tend to have higher grade than men who drop.

It is therefore important to understand whether, for the minority of women who engage in STEM disciplines in higher education, there is some evidence that their achievements are measurably different or are penalized by the curricular design, the bias existing in male-dominated fields, or by the teaching environment itself. As a first step in attempting to gain some insight into this question, we study in this paper the existence of differences in academic performance between men and women in a typical computer networking course, taught at an introductory level.

Ours is a cross-sectional study of a single group of students at the University of Vigo (Spain), pursuing a bachelor of engineering. The group has a prototypical composition wherein only 20% of the cohort are female students, which is in agreement with other engineering undergraduate degrees in our university. We have not designed or executed specific assessment tasks for discovering possible differences in performance. Quite on the contrary, our analysis is performed ex-post on the grades attained in the different learning tasks proposed to the students. We intend to uncover any difference in achievements obtained by men and women, if any, and a possible explanation to those.

The organization of the manuscript is as follows. Some related articles are collected and summarized in Section 2. Section 3 describes the content and the methodology employed in the course under study. Section 4 contains the results of different tests applied to the data to quantify the influence of gender on course engagement and its impact on learning results. Finally, some conclusions are included in Section 5.

Sousa-Vieira, M., López-Ardao, J. and Fernández-Veiga, M.
Study of Impact of Gender on Engagement and Performance of Engineering Students.
DOI: 10.5220/0012763300003756
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 13th International Conference on Data Science, Technology and Applications (DATA 2024)*, pages 358-367
ISBN: 978-989-758-707-8; ISSN: 2184-285X
Proceedings Copyright © 2024 by SCITEPRESS – Science and Technology Publications, Lda.

2 RELATED WORK

Gender studies in higher education are mainly related to the gender gap in enrollment in STEM degrees and, to a lesser extent, to the analogies and differences in engagement and performance of both groups of students, men and women, in higher studies in general or in certain subjects in particular. In this Section, we review a sample of some representative works in each field.

Related to the first field of study, the objective of the research conducted in (Borsotti, 2018) was to understand the socio-cultural factors affecting women's decisions regarding enrollment in a bachelor of Software Development in Denmark. The results were aligned with other recent studies, highlighting that primary obstacles to female involvement in computer science education in Denmark are a combination of gender differences in previous familiarity and experience with the field of computer science, pervasive gender biases regarding aptitude for pursuing careers in computer science and stereotyped perceptions associated with software development professions. Research findings of the study promoted the development of local interventions, resulting in the recruitment of twice as many women in one year. In (Ehrlinger et al., 2018), college students of the United States were asked to provide descriptions and trait ratings for the typical computer scientist or engineer. Authors found that women tended to offer ratings more aligned with stereotypes regarding the intellectual traits associated with prototypes in computer science and engineering compared to men. Additionally, women perceived themselves as less similar to the prototypes than men did. Furthermore, the gender disparities observed in perceptions of prototypes played a mediating role in women's tendency to express lower interest in these fields compared to men. The study described in (Finzel et al., 2018) introduced the makeIT mentoring program, which aimed to inspire female high school students in Germany to explore computer science as a potential career path. The program's design drew upon insights into the barriers that discourage girls from pursuing STEM fields, particularly computer science, as well as findings on the impact of mentoring on psychosocial factors. In (Olmedo-Torre et al., 2018), quantitative and qualitative investigations were conducted in Spain to ascertain the existence and nature of stereotypes regarding women pursuing STEM studies, particularly in fields such as computer science, communications, electrical and electronic engineering. The quantitative findings indicated the presence of stereotypes with variations between women pursuing technological and engineering degrees and women in other STEM fields. The qualitative analysis revealed that women perceive social stereotypes and their immediate surroundings as significant factors contributing to their underrepresentation in STEM studies.

The paper (Volkel et al., 2018) provided an examination of the contemporary factors driving individuals to choose computer science as their field of study in Germany, with a specific focus on gender dynamics. The findings confirm that women and men are motivated by different reasons and hold different expectations regarding computer science. Moreover, in some cases they evaluate their strengths and abilities in different ways. Both agreed that the primary motivator is interest in the contents, but women tended to rate their interest lower than men. This tendency stemmed from women's perception of the prototypical computer scientist as possessing extensive expertise in programming and technology, leading them to undervalue their own interest when comparing themselves to this idealized image. In (Ballatore et al., 2020) authors presented the conclusions of a collaborative project involving six universities across Europe (in Germany, Sweden, Italy, Belgium, Ireland and Portugal) which aimed to explore gender disparities in students' self-perception regarding their career choices. The conclusions revealed that women are less likely to be influenced by friends' opinions compared to men students and place greater emphasis on parental encouragement. Moreover, women expect more classmates to have a lot of hands-on experience, are more likely to have doubts about their mathematical abilities relative to their peers and are more likely to perceive engineering difficult than men. The study (López-Iñesta et al., 2020) presented the objectives and main activities of a pilot program in Spain to promote STEM careers, focusing on increasing and retained the number of women choosing these studies. Preliminary results demonstrated the statistically significant increase in the proportion of women enrolled in Computer Science, Electronics and Telematics in the university where the program was executed. Finally, the paper (Davila et al., 2022) proposed a theoretical framework to address the issue of low participation of women in STEM fields in Spain. The framework proposed strategies to raise awareness and promote women involvement in scientific and technology careers. Results suggested that it should be possible to narrow the gender gap in career choices among young students by implementing a series of targeted activities directed at individual students, their families and peers, schools and society at large, aimed at changing longstanding habits that have historically deterred women from pursuing STEM studies.

Surprisingly, there are not many studies concerning the analogies and differences in engagement and performance of men and women in STEM higher education. In this sense, our study constitutes a new contribution to the state of the art on this subject. Following, we provide a representative sample of the few articles we have found. In (Caspi et al., 2008) authors examined gender disparities in participation between face-to-face and web-based classroom discussions, by comparing the actual participation ratio with the attendance (or login) ratio of men to women in the open university in Israel. The research uncovered that men tended to speak disproportionately more in face-to-face settings, whereas women exhibited a higher proportion of message posting in web-based conferences, suggesting that either women have a preference for written communication over verbal communication, or women prefer written communication more than men. The purpose of (Ammons and Brooks, 2011) was to investigate gender variations in student evaluations of individual contributions to collaborative projects within an undergraduate business course in the United States. The study concluded that women receive higher ratings by both men and women evaluators. Moreover, evaluators of the same gender as the students being assessed provided more comprehensive feedback, including both positive and negative comments, compared to those of the opposite gender. In addition, women tended to both receive and provide more open-ended feedback than men, and this feedback tended to be positive reinforcement. Interestingly, no gender disparities were observed in self-assessment ratings. The work in (Tucker, 2014) considered the possibility of gender bias in peer ratings for contributions to team assignments, as measured by an online self-and-peer assessment tool over participants from two universities of Australia, enrolled in four different degree programs (business and law, environments, design and construction management). The study concluded that there was an absence of gender bias in the peer ratings and, notably, women received significantly higher ratings than men. The study (Martinho et al., 2015) investigated gender disparities in cooperation and competitiveness in a first-year chemistry course in Portugal. The findings supported the notion that women students tended to exhibit more cooperative behavior and less competitiveness compared to their male counterparts. Using social network analysis, authors of (Grunspan et al., 2016) explored how gender influences the confidence that college-level biology students in the United States have in each other's mastery of course content. The results revealed than men are more likely than women to be nominated by their peers as knowledgeable about the course content. This was attributed to men's tendency to over-nominate their men peers relative to their actual performance. Women, in contrast, tend to nominate peers based on performance rather than gender. This disparity in social dynamics can be one of the reasons for women to leave natural and physical sciences studies in greater proportion than their male counterparts.

In (Chopra et al., 2019) a data-intensive approach was employed to investigate gender differences in STEM fields in the United States. Using deep learning, text mining and statistical methodologies, the study analyzed various academic datasets, including undergraduate admission records, descriptions of cooperative job and data on student entrepreneurship. Results shown that women exhibit distinct motivations compared to men when applying to engineering programs, and that women tend to occupy slightly different cooperative roles during their undergraduate studies and show less inclination towards entrepreneurial endeavors. The purpose of the study (Noroozi et al., 2020) was to explore gender-based distinctions in argumentative feedback quality, essay writing and content learning in an online educational setting within a biotechnology degree program in the Netherlands. The findings revealed disparities between women and men in terms of the quality of their argumentative feedback, higher for women. Although all students exhibited improvements in essay writing quality and knowledge content from pre-test to post-test, the study found no significant differences in these improvements between women and men. In (Kazem et al., 2021) researchers delved into gender differences in engagement and self-regulation within a constructivism learning framework, coupled with a learning analytics environment and implemented in a teaching skills course in a public university in Iran. The findings indicated that women students exhibited higher levels of engagement and self-regulation compared to their men counterparts, with scores increasing from pre-test to post-test. This suggest that women students found the online environment to be more engaging than men did. Finally, the article (Ibarra-Vazquez et al., 2023) aimed to investigate the performance of machine learning models in classifying Mexican students by gender based on their perceived competency in complex thinking. The results validated the hypothesis than the four machine learning models tested were capable of detecting sufficient differences in the data to accurately classify students by gender during both the training and testing phases.

3 EDUCATIONAL ENVIRONMENT

Our study have focused on the 2020/2021 edition of a course on Computer Networks tailored for second-year undergraduates enrolled in the Telecommunications Technologies Engineering degree (Sousa-Vieira et al., 2023). Spanning 14 weeks form January to May, the course comprises various classroom activities structured as follow:

- Lectures: These sessions blend the presentation of fundamental concepts, techniques and algorithms with practical exercises and discussion of theoretical or practical questions.
- Laboratory sessions: Here, students engaged in designing and analyzing diverse network scenarios with different protocols, using either real or simulated networking equipment. Additionally, some sessions involve students undertaking programming projects.

To support these activities, a Moodle platform was utilized, facilitating communication between students and teachers regarding the covered topics. In order to foster self-learning and collaborative work, a range of activities are planned and carried out in the platform. Throughout the edition analyzed in this research, the following online activities were proposed:

- Homework tasks: Designed to be completed before the in-class or the laboratory sessions, these tasks encourage students to prepare some of the material in advance.
- Quizzes: Proposed before the midterm exams, these quizzes provide students with opportunities for self-training.
- Collaborative participation in forums: Several forums were established on Moodle, allowing students to post questions related to the course organization, lectures content, laboratory sessions and programming assignments.
- Optional activities: These include games, peers assessments and other supplementary tasks.

Students have the opportunity to earn different kinds of recognition through their engagement in these activities.

The performance in tasks and quizzes, including their peers assessments, is quantified in merit points, contributing to the total score gained for accomplishment of these activities. Extra merit points can be earned by completing the optional activities, serving as a means to compensate for low scores or late submissions of some of the tasks or quizzes. Well-done peers assessments and outstanding scores in tasks and quizzes are rewarded with coins and badges.

Active participation in forums, where students can address doubts, share resources and engage in discussions, is also valued. Accumulating points and votes, granted by teachers or classmates, elevates the student's karma level, depending on different factors that take into account the quality of the student's actions and the comparison with that of his/her classmates. This can result in the acquisition of coins.

Engagement within the virtual classroom is incentivized through the automatic scoring of various activities conducted on the platform. These activities encompass tasks such as resource posting or viewing, initiating new threads of discussions, responding to posts, completing assignments, and so forth. Termed as experience points, these points are organized into levels and are granted within a regulated framework, with predefined maximum values and frequency, overseen by the teachers. As students progress through levels, they also get coins.

Students have access to their accumulated merit points, karma level, and experience points and level at any given time. Additionally, they can check their positions in global rankings and the average values. Occasionally, top-performing students of a ranking may be publicly acknowledged within the group.

Once coins become available in the warehouse, students have the option to collect and store them in their inventory. These accumulated coins can later be exchanged for various benefits helpful in passing the subject. The conversion options include:

- Exchanging 8/16 coins for additional 10/20 min of time in the final exam.
- Exchanging 12/24 coins for one/two pages of notes for the final exam.
- Exchanging 8 coins for 5 bonus merit points up to a maximum of 25 points.

Students may pass the course after a single final examination covering all the material (provided the programming assignment meets the minimum requirements), but they are encouraged to follow the continuous assessment. The continuous assessment allows two modalities, A and B. The final exam represents the 50% of the final grade, but the rest is split as follows: 30% in modality A and 20% in modality B from the midterm exams, 10% from the programming project and 10% (only in modality B) coming out from the merit points obtained by accomplishing the online activities (tasks, quizzes and optional activities) described previously, devised as a tool to increase the level of participation. It is evident that the students that follow the modality B of the continuous

Table 1: Basic statistical parameters and p-values of the Kolmogorov-Smirnov test. Quizzes for self-training.

Q1	Men			Women		
	$\hat{\mu}$	$\hat{\sigma}^2$	p-KS	$\hat{\mu}$	$\hat{\sigma}^2$	p-KS
Grade	7.3231	13.3856	0.0461	8.4671	18.8678	0.5331
Completion Time	43.8381	148.8101	0.1444	47.9143	114.7277	0.3602
Correct answers	8.3333	8.8974	0.0096	9.0571	11.2319	0.3864
Failed answers	4.6762	7.0864	0.0041	3.7714	8.1815	0.2188
Unanswered questions	1.1619	2.6562	$< 10^{-8}$	1.1428	2.3613	0.0195
Q2	**Men**			**Women**		
	$\hat{\mu}$	$\hat{\sigma}^2$	p-KS	$\hat{\mu}$	$\hat{\sigma}^2$	p-KS
Grade	7.2454	18.5467	0.0813	10.1258	12.6458	0.6431
Completion Time	44.4875	173.5404	0.1028	49.0312	90.2893	0.5155
Correct answers	10.8571	16.2793	0.0448	13.1562	11.8135	0.5419
Failed answers	5.5714	11.9143	0.0232	4.0312	10.2893	0.0907
Unanswered questions	1.5714	4.6254	$< 10^{-4}$	0.8751	1.9838	0.0021
Q3	**Men**			**Women**		
	$\hat{\mu}$	$\hat{\sigma}^2$	p-KS	$\hat{\mu}$	$\hat{\sigma}^2$	p-KS
Grade	9.0089	19.7911	0.0931	11.7014	14.3499	0.1161
Completion Time	40.3243	176.6605	0.7191	37.9655	167.4631	0.8064
Correct answers	10.2432	13.2003	0.0587	12.4137	9.9655	0.0632
Failed answers	3.4729	8.3074	0.0576	1.7241	5.6354	0.0077
Unanswered questions	1.2837	3.1375	$< 10^{-4}$	0.9311	2.4951	0.0022

Table 2: p-values of the Fligner–Killeen and the Mann-Whitney-Wilcoxon tests. Quizzes for self-training.

	p-FK			p-MWW		
	Q1	Q2	Q3	Q1	Q2	Q3
Grade	0.2433	0.0938	0.1519	0.0974	0.009	0.0037
Completion Time	0.1184	0.0611	0.7392	0.1056	0.1901	0.4071
Correct answer	0.2415	0.0712	0.1461	0.1895	0.0054	0.0035
Failed answer	0.7827	0.0794	0.0722	0.0627	0.0271	0.0014
Unanswered question	0.6767	0.0661	0.0511	0.8275	0.1341	0.2733

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

Table 3: Basic statistical parameters and p-values of the Kolmogorov-Smirnov test. Middle term exams.

EQ1	Men			Women		
	$\hat{\mu}$	$\hat{\sigma}^2$	p-KS	$\hat{\mu}$	$\hat{\sigma}^2$	p-KS
Grade	4.6887	3.9064	0.9979	3.5108	6.5244	0.7606
Correct answer	7.3534	4.5435	0.0625	6.0294	8.2718	0.4091
Failed answer	4.6637	4.3121	0.0053	5.2353	5.2762	0.0362
Unanswered question	3.9741	4.7384	0.0322	4.7352	5.3521	0.1178
EQ2	**Men**			**Women**		
	$\hat{\mu}$	$\hat{\sigma}^2$	p-KS	$\hat{\mu}$	$\hat{\sigma}^2$	p-KS
Grade	4.6486	6.5485	0.5773	5.1691	8.2966	0.6531
Correct answer	8.7451	12.8254	0.4153	9.2258	16.1139	0.9102
Failed answer	5.0981	9.4161	0.3517	3.3548	9.7032	0.0881
Unanswered question	4.0588	6.2341	0.7559	5.4193	9.7849	0.6146

Table 4: p-values of the Fligner–Killeen and the Mann-Whitney-Wilcoxon tests. Middle term exams.

	p-FK		p-MWW	
	EQ1	EQ2	EQ1	EQ2
Grade	0.1779	0.4614	0.0019	0.3132
Correct answer	0.1618	0.5258	0.0014	0.6021
Failed answer	0.2951	0.6225	0.1579	0.0016
Unanswered question	0.6221	0.2072	0.0657	0.0233

assessment can gain more advantages from the gamification strategy.

To finish the description, among all students enrolled in this course, 102 students (71% of the enrolled women and 63% of the enrolled men) did not drop out (i.e. they attended the final exam). Among these 102 students, of the 101 students who followed the continuous assessment (14 in modality A (2 women and 12 men) and 87 in modality B (23 women and 64 men), 73 finally passed the course (6 (43%) in modality A (all men), and 67 (77%) in modality B (15 women and 52 men). The student that followed the single final exam modality was not able to pass. In conclusion, related to gender, 43% of the enrolled women (60% of the women that did not drop out) and 47% of the enrolled men (75% of the men that did not drop out) passed the course.

In the following sections we try to answer the following research question:

RQ. What was the impact of gender on students engagement and performance?

4 ANALYSIS OF THE DATASETS

In this section, we show the results of different tests to quantify the influence of gender on course engagement and its impact on learning results.

4.1 Quizzes for Self-Training

Firstly, we have analysed different metrics (grade, completion time and number of correct answers, failed answers and unanswered questions) related to the performance of students along the term in quizzes (Q1, Q2 and Q3) for self-training before the middle term exams and the final exam. Any student can take these tests but they have only influence in the final grade for those who follow the modality B of the continuous assessment. Q1 and Q3 had 15 questions and Q2 had 18 questions. In all cases, the maximum response time was 60 minutes. Each correct answer adds one point and each failed answer subtracts $\frac{1}{N-1}$ points, being N the number of possible answers (only one is correct). In Table 1 we show the sample mean value and variance of several metrics related to these quizzes, separated by gender. We can see that, on average, women achieve best grades, expend slightly more time in each quiz and answer more questions.

In order to check if there are significant differences among the performance of both groups of students, men and women, we have applied different statistical tests. Firstly, the Kolmogorov–Smirnov test of normality. In Table 1 the p-values obtained confirm that not for all the cases we cannot conclude that the metrics do not follow a normal distribution ($p > 0.05$). For this reason, in a second step, we have applied the Fligner–Killeen (FS) test of comparison of variances to the metrics related to Q1, Q2 and Q3 of both groups of students. According to Table 2, the p-values obtained confirm that for all pairs of samples there is no evidence against equality of variances ($p > 0.05$). As a final check, we employed the Mann–Whitney–Wilcoxon (MWW) test to assess the homogeneity/heterogeneity of both groups with respect to the metrics in each quiz, because this test does not require normality, although it does require equality of variances. Results are listed in Table 2 too, and now the p-values imply that for some samples we cannot discard that there are no differences between both groups ($p > 0.05$), whereas for others (grades and correct and failed answers in Q2 and Q3), the distributions of the metrics of both groups are different ($p \leq 0.05$).

4.2 Middle Term Exams

In a similar fashion, we also analysed similar metrics (grade and number of correct answers, failed answers and unanswered questions) related to the performance of students in the middle term exams (17^{th} of March and 5^{th} of May), consisting of two quizzes (EQ1 and EQ2). These quizzes are only accounted in the final grade for students who follow the continuous assessment. EQ1 consisted of 20 questions, and EQ2 had 18 questions. In both cases, the maximum response time was 60 minutes. Again, each correct answer adds one point and each failed answer subtracts $\frac{1}{N-1}$ points, where N denotes the number of possible answers (only one is correct). The basic statistical analysis is shown in Table 3, which includes the sample mean value and the variance of several metrics related to these quizzes, separated by gender. In this case, on average, men achieve best grades in EQ1 and women in EQ2 but, in both cases, men answer more questions.

As for the statistical significance of the differences between genders, the same tests (K-S, FK, MWW) used in the previous Section were applied. The K-S test (included in Table 3) has p-values showing that, for some of the metrics, the hypothesis of normality cannot be rejected ($p > 0.05$). Nevertheless, for the FK test (Table 4) the conclusion is that for all pairs of samples there is no evidence against equality of variances ($p > 0.05$). The homogeneity/heterogeneity of both groups with respect to the metrics in each exam is specifically analyzed through the MWW test, and here our results give evidence that for some samples it is not possible to discard that there are no differences between both groups ($p > 0.05$), but for others (grades and correct answers in EQ1 and failed answers in EQ2), the distributions of the metrics of both groups are actually different ($p \leq 0.05$).

4.3 Homework Tasks

In the next step, we have analysed the performance (grade and peers assessment) of students in homework tasks (T0, T1, T2 and T3) along the term. T0

Table 5: Basic statistical parameters and p-values of the Kolmogorov-Smirnov test. Homework tasks.

T0	Men			Women		
	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS
Grade	15.8746	35.3921	$< 10^{-4}$	17.9562	11.5091	0.0178
Peers assessment	5.3041	0.7346	0.0001	5.4532	0.4535	0.1755
T1	**Men**			**Women**		
	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS
Grade	5.2758	3.0607	$< 10^{-4}$	5.5506	3.8391	0.0393
Peers assessment	2.7193	0.1513	$< 10^{-8}$	2.7061	0.1873	0.0007
T2	**Men**			**Women**		
	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS
Grade	6.9389	4.4761	0.0009	7.7651	7.5369	0.0201
Peers assessment	2.8389	0.0551	0.0006	2.8593	0.0581	0.0182
T3	**Men**			**Women**		
	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS
Grade	5.5243	5.4152	0.0003	5.9119	3.5869	0.0546
Peers assessment	1.8082	0.0691	0.0011	1.7746	0.0532	0.1603

Table 6: p-values of the Fligner–Killeen and the Mann-Whitney-Wilcoxon tests. Homework tasks.

	p-FK			
	T0	T1	T2	T3
Grade	0.0753	0.6219	0.0837	0.9856
Peers assessment	0.4046	0.1361	0.9389	0.4072
	p-MWW			
	T0	T1	T2	T3
Grade	0.2653	0.3573	0.9642	0.6801
Peers assessment	0.5498	0.6907	0.4811	0.3161

was a small programming exercise for training before the programming project. T1 and T3 were exercises about the first and last part of the theory and T2 was a task of configuration and traffic analysis to be done with a network simulator. Again, any student can do these tasks but they have only influence in the final grade for those who follow the modality B of the continuous assessment.

For this type of homework assignments, our analysis produced the results listed in Tables 5 and 6 for the basic statistics and K-S test, and the FK and MWW tests, respectively. The first Table shows that, on average, women achieve slightly best grades in each task. Related to the quality of peers assessment, results of both groups are very similar. For the hypothesis of normal distributions, Table 5 contains p-values showing that it is not possible to reject the hypothesis for all the tested cases, since $p > 0.05$. At the same time, we did not find evidence that the pairs of samples have different variances (FK test, Table 6), and similarly that it is not possible to discard that no differences exist between both groups (MWW test, Table 6).

4.4 Programming Project and Final Exam

Finally, we conducted the analysis of the students' performance in the programming project and in the final exam (abbreviated as PRO and FE, respectively). Tables 7 and 8 lists our numerical results for these learning tasks. A first observation is that, on average, men get slightly best grades in the project and in the final exam. As in the previous cases, it was not possible to reject the hypothesis of normal distribution for all the subgroups, and it was not possible either to find evidence against the distributions having equal variance. Regarding the homogeneity of both groups, we can not discard that there are no differences between both them ($p > 0.05$).

4.5 Participation in Forums

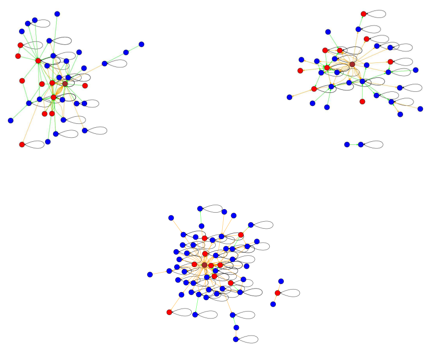

Figure 1: Forums activity graphs. Lessons (top-left), programming (top-right) and organization (bottom).

We have recorded the events that took place in each forum (lessons forum for queries related to the concepts of the subject, programming forum for queries

Table 7: Basic statistical parameters and *p*-values of the Kolmogorov-Smirnov test. Programming project and final exam.

	Men			Women		
	$\hat{\mu}$	$\hat{\sigma^2}$	*p*-KS	$\hat{\mu}$	$\hat{\sigma^2}$	*p*-KS
Grade PRO	5.7621	5.2297	$< 10^{-6}$	5.0521	8.5488	0.0031
Grade FE	4.3827	2.6637	0.7994	4.0012	3.9215	0.2932

Table 8: *p*-values of the Fligner–Killeen and the Mann-Whitney-Wilcoxon tests. Programming project and final exam.

	p-FK	*p*-MWW
Grade PRO	0.4666	0.9351
Grade EF	0.0599	0.2401

related to the programming tasks and organization forum for queries related to the organization of the course): users who initiated new threads, users who replied, and the ratings they received. This information is visualize through graphs, where each user is represented as a node. Connections between nodes (edges) indicate replies, that is, two users are connected by an edge if one has given a reply to an entry posted by the other. Moreover, a self-edge denotes a new thread initiated by this user. Green edges highlight the best replies, as determined by the owner of the question and/or the teachers. An illustration of the graphs related to each forum is given in Figure 1. Students are differentiated by color; nodes in blue represent men and nodes in red represent women. The node in brown corresponds to the instructors.

In comparison with the percentage of women in the course, we can see a higher ratio of women in the lessons forum (11 of 36 participants) and in the programming forum (9 of 34 participants). Nevertheless, the presence of women in the organization forum is low in comparison with men (10 of 56 participants).

Lastly, we have considered the type of participation (number of new threads, number of replies and points received for replies) in each forum (LF, PF and OF) of women and men (taking into account all the students that followed the continuous assessment). In Table 9 we can see that, on average, women achieve greater values in all metrics. For the hypothesis of normal distributions, the *p*-values show that, for all the samples, the hypothesis of normal distributions can be rejected. For the FK test (Table 10) the conclusion is that for all pairs of samples there is no evidence against equality of variances ($p > 0.05$). The homogeneity/heterogeneity of both groups with respect to the metrics in each forum is analyzed through the MWW test (Table 10), and here our results give evidence that for all the samples it is not possible to discard that there are no differences between both groups ($p > 0.05$).

5 CONCLUSIONS

After the statistical analysis reported in this paper, we can conclude that, according to our sample results, there is no statistically significant gaps between men and women in our focused study group. We just observed minor differences in each individual task proposed thoroughout the academic semester, with marginally better results for women in homeworks and slightly better outcomes for men in programming and in written examinations. Moreover, we measured some positive differences regarding the participation of women in social activities.

A limitation of this study is its cross-sectional nature, and the small size of the sample (just a little over 100 students), so our results cannot be generalized without caution. However, we believe that our analysis suggests that women do not suffer from bias in expectations in their work in this course, and have indistinguishable learning outcomes compared to their male classmates.

To gain more understanding about whether women exhibit or not more/less collaboration/competitiveness than men, we plan to apply other techniques of analysis on our data, since the course includes activities in an online site and allows for the spontaneous creation of a social network of students. We also have plans for extending this study to more cohorts, thus building up a longitudinal study with further information on the evolution of the variables.

REFERENCES

Alexandru, M., Whitcomb, K., and Singh, C. (2022). Gender inequities throughout stem: Compared to men, women with significantly higher grades drop stem majors. *Journal of College Science Teaching*, 51(3).

Ammons, J.-L. and Brooks, C.-M. (2011). An empirical study of gender issues in assessments using peer and self evaluations. *Academy of Educational Leadership Journal*, 15:49–62.

Ballatore, M.-G., Borger, J., Misiewicz, J., and Tabacco, A. (2020). Anna tool: A way to connect future and past students in stem. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 15(4):344–351.

Table 9: Basic statistical parameters and p-values of the Kolmogorov-Smirnov test. Participation in forums.

LF	Men			Women		
	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS
# new threads	0.0975	0.1051	$< 10^{-15}$	0.7428	0.7261	$< 10^{-6}$
# replies	0.2764	0.6442	$< 10^{-15}$	0.8001	3.2823	$< 10^{-4}$
Points replies	0.5609	2.4122	$< 10^{-15}$	1.3428	7.7025	$< 10^{-5}$
PF	**Men**			**Women**		
	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS
# new threads	0.1382	0.1692	$< 10^{-15}$	0.4285	1.2521	$< 10^{-5}$
# replies	0.2276	0.8166	$< 10^{-15}$	0.2857	0.4453	$< 10^{-6}$
Points replies	0.3821	2.3855	$< 10^{-15}$	0.4285	0.8403	$< 10^{-6}$
OF	**Men**			**Women**		
	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS	$\hat{\mu}$	$\hat{\sigma^2}$	p-KS
# new threads	0.3171	0.3306	$< 10^{-15}$	0.5143	2.1394	$< 10^{-4}$
# replies	0.2113	0.2664	$< 10^{-15}$	0.4571	1.5495	$< 10^{-6}$
Points replies	0.2195	0.3694	$< 10^{-15}$	0.6285	2.5932	$< 10^{-6}$

Table 10: p-values of the Fligner–Killeen and the Mann-Whitney-Wilcoxon tests. Participation in forums.

	p-FK			p-MWW		
	LF	PF	OF	LF	PF	OF
# new threads	0.1412	0.0858	0.8225	0.2981	0.1584	0.8233
# replies	0.0541	0.3457	0.4974	0.1071	0.3609	0.8653
Points replies	0.0811	0.3621	0.2382	0.1636	0.2992	0.5885

Borsotti, V. (2018). Barriers to gender diversity in software development education: Actionable insights from a danish case study. In *ICSE'18, 40th ACM/IEEE International Conference on Software Engineering*.

Caspi, A., Chajut, E., and Saporta, K. (2008). Participation in class and in online discussions: Gender differences. *Computers & Education*, 50:718–724.

Chopra, S., Mirsafian, M., Khan, A., and Golab, L. (2019). Gender differences in science and engineering: A data mining approach. In *EDBT/ICDT'19, 22th International Joint Conference on Extending Database Technology and Theory*.

Davila, E., Albahari, A., Díaz, S., and de Freitas, E.-C. (2022). Science and technology as feminine: Raising awareness about and reducing the gender gap in stem careers. *Journal of Gender Studies*, 31(4):1–14.

Ehrlinger, J., Plant, E.-A., Hartwig, M.-K., Vossen, J.-J., Columb, C.-J., and Brewer, L.-E. (2018). Do gender differences in perceived prototypical computer scientists and engineers contribute to gender gaps in computer science and engineering? *Sex Roles*, 78:40–51.

Finzel, B., Deininger, H., and Schmid, U. (2018). From beliefs to intention: Mentoring as an approach to motivate female high school students to enrol in computer science studies. In *GenderIT'18, 4th Conference on Gender & IT*.

Grunspan, D.-Z., Eddy, S.-L., Brownell, S.-E., Wiggins, B.-L., Crowe, A.-J., and Goodreau, S.-M. (2016). Males under-estimate academic performance of their female peers in undergraduate biology classrooms. *PloS ONE*, 11(2):1–16.

Hundschell, A., Razinskas, S., Backmann, J., and Hoegl, M. (2021). The effects of diversity on creativity: A literature review and synthesis. *Applied Psychology*, 71(4):1598–1634.

Ibarra-Vazquez, G., Ramírez-Montoya, M.-S., and Terashima, H. (2023). Gender prediction based on university students' complex thinking competency: An analysis from machine learning approaches. *Education and Information Technologies*, 28(6):1–19.

Kazem, S., Noroozi, O., and Akhteh, M.-P. (2021). Gender differences in engagement and self-regulation in an online constructivist learning design and learning analytics environment. In *ICSEE'21, 2nd International Conference on Studies in Education and Social Sciences*.

López-Iñesta, E., Botella, C., Rueda, S., Forte, A., and Marzal, P. (2020). Towards breaking the gender gap in science, technology, engineering and mathematics. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 15(3):233–241.

Martinho, M., Albergaria-Almeida, P., and Teixeira-Dias, J. (2015). Cooperation and competitiveness in higher education science: Does gender matter? *Procedia - Social and Behavioral Sciences*, 191:554–558.

Noroozi, O., Hatamia, J., Bayatc, A., Ginkel, S., Biemans, H., and Mulde, M. (2020). Students' online argumentative peer feedback, essay writing and content learning: Does gender matter? *Interactive Learning Environments*, 28(6):698–712.

Olmedo-Torre, N., Sánchez-Carracedo, F., Salán-Ballesteros, M.-N., López, D., Pérez-Poch, A., and López-Beltrán, M. (2018). Do female motives for enrolling vary according to stem profile? *IEEE Transactions on Education*, 61(4):289–297.

Sousa-Vieira, M.-E., López-Ardao, J.-C., Fernández-Veiga, M., and Rodríguez-Rubio, R.-F. (2023). Study of the impact of social learning and gamification methodologies on learning results in higher education. *Computers Applications in Engineering Education*, 31(1):131–153.

Tucker, R. (2014). Sex does not matter: Gender bias and gender differences in peer assessments of contributions to group work. *Assessment & Evaluation in Higher Education*, 39(3):293–309.

Volkel, S.-T., Wilkowska, W., and Ziefle, M. (2018). Gender-specific motivation and expectations toward computer science. In *GenderIT'18,* 4^{th} *Conference on Gender & IT*.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS