Understanding Student Evaluation of Teaching in Computer Science Courses

Zixuan Wang
zwan843@aucklanduni.ac.nz
University of Auckland
Auckland, New Zealand

Paul Denny
University of Auckland
Auckland, New Zealand
paul@cs.auckland.ac.nz

Juho Leinonen
University of Auckland
Auckland, New Zealand
juho.leinonen@auckland.ac.nz

Andrew Luxton-Reilly
University of Auckland
Auckland, New Zealand
andrew@cs.auckland.ac.nz

ABSTRACT

Understanding student perceptions in higher education is vital for optimizing teaching and learning practices. This research explores the relationship between course characteristics, Student Evaluation of Teaching, and disciplinary differences, with a particular focus on Computer Science courses. Analyzing data from the second half of the 2022 semester at one university, the study investigates the impact of course level, type, and size on student evaluation scores. Additionally, it compares Computer Science courses to other disciplines, revealing key differences in student satisfaction and perceptions. Findings indicate that second-year courses received lower ratings, and theoretical courses in online formats received higher satisfaction than programming courses. Smaller course sizes correlated with higher scores across multiple aspects. However, Computer Science courses scored lower overall and in crucial areas compared to other disciplines, highlighting the need for tailored teaching strategies. This research underscores the importance of continuous assessment and adaptation in higher education to foster positive learning environments and improve student experiences.

CCS CONCEPTS

• **Social and professional topics → Computing education**.

KEYWORDS

Student Evaluation of Teaching, Student Feedback, Student Engagement

ACM Reference Format:
Zixuan Wang, Paul Denny, Juho Leinonen, and Andrew Luxton-Reilly. 2023. Understanding Student Evaluation of Teaching in Computer Science Courses. In *16th Annual ACM India Compute Conference (COMPUTE '23), December 09–11, 2023, Hyderabad, India.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3627217.3627220

COMPUTE '23, December 09–11, 2023, Hyderabad, India

ACM ISBN 979-8-4007-0840-4/23/12…$15.00
https://doi.org/10.1145/3627217.3627220

1 INTRODUCTION

Student evaluations of teaching are questionnaires completed by students to provide feedback on their experiences and perceptions of teaching staff and course effectiveness. Students are typically invited to rate their teachers' performance on Likert-scale questions and are then asked to provide anonymous narrative comments. Understanding the factors that influence student evaluation scores and the differences observed across various disciplines can offer valuable insights for educators, institutions, and policymakers seeking to optimize the educational environment [1].

We analysed the student evaluation dataset (SET) for the second half of the 2022 semester at one university and conducted a study. We sought to uncover the complex relationship between course characteristics, SET indicators, and disciplinary differences. We explored the impact of different course characteristics (including course level, course size, and discipline) on students' scores on multiple aspects, such as collaboration, clarity, and overall instructional quality. Uncovering the relationships between these factors has the potential to reveal essential considerations for educators when designing programs that resonate with students and foster positive learning experiences.

Computer Science deserves particular attention as a subject that has seen significant growth in recent years. It has been shown that there are significant differences between computer science students and other students in terms of student engagement [4]. We have also analysed in-depth comparisons between computer science courses and other disciplines with regard to specific aspects of SET scores. Our aim was to discover how these differences relate to various characteristics of computer science courses, such as the type and stage of the course. By identifying these differences, we may be able to inform the development of more appropriate teaching strategies in computer science and other areas. Hence, our research questions are:

RQ 1: How do different course features (such as course level, course size, and discipline) impact the scores on various components of student evaluation of teaching?

RQ 2: What specific aspects of student evaluation scores differ between Computer Science courses and other disciplines, and how do these differences relate to various features of Computer Science courses?

2 RELATED WORK

SET is a survey instrument that collects student feedback on teaching staff and courses. Despite their use for teaching improvement, research indicates that student evaluation scores are not valid indicators of teaching competence [1, 5, 6]. However, they continue to be used by University administrative bodies [4].

Computer Science and Software Engineering courses have some special characteristics compared to other academic courses. The teaching of Computer Science courses includes a lot of technical content and teaching the use of software tools, so students may be more concerned about the practicality of the content and whether it will improve their programming skills [4].

Ivins et al. conducted a study on first-year and fourth-year Software Engineering students' perceptions [2]. First-year students prioritized programming and management skills, while fourth-year students rated requirements gathering, analysis, design, and testing skills higher. First-year students also emphasized mathematical skills, while fourth-year students valued soft skills like communication, leadership, and teamwork. These findings highlighted misconceptions and stereotypes surrounding Software Engineering and suggested the need for curriculum adjustments.

Knutas et al. investigated biases in student evaluation of teaching for Software Engineering courses [3]. Analyzing 1295 student ratings from 46 courses, they found that programming courses received higher student evaluation scores than those focused on Software Engineering processes, models, and methods. These biases can impact students' responses. The study emphasized the need to consider biases when analyzing student evaluation of teaching results and use multiple indicators to evaluate teaching effectiveness. The analysis showed that courses emphasizing software construction and programming generally received higher student ratings [3]. This confirmed the existence of biases observed in prior Software Engineering education studies.

Morgan et al. used an analysis of the Computer Science education literature and interviews with Computer Science academics to examine the differences in student engagement between computing students and other students [4]. They compared and contrasted the focus of previous computing education research and the engagement measures used in various tools. The study found that there were significant differences in student engagement between computing students and other students. They found there were deficiencies in the perceptions of computing course instructors regarding the concept of student engagement. The pedagogical practices of computing teaching and learning have limitations in promoting student engagement. Computer Science students often prefer individual learning and independent thinking, with less emphasis on teamwork and communication. This may be related to the large class sizes and lack of opportunities for interaction and individualised learning that is prevalent in computer science education. Teachers may also face constraints in their teaching in terms of curriculum and classroom resources, which further limit their teaching practices in terms of student engagement. In addition, the study found that computing students had some difficulties in collaborating, communicating and forming learning communities with other students, which may be related to the professional characteristics of computer science and the culture of the discipline. They suggest that the differences in student engagement between computing students and other students may be the result of a combination of factors.

3 QUANTITATIVE SET DATA ANALYSIS

We conducted an analysis of the SET rating data for 272 Science courses in the second semester of 2022 at the University of Auckland. A total of 4,759 students participated in the evaluation, including 3,895 on-campus students and 864 online learning students. The courses were divided into four levels: Stages 1, 2, 3, and 7 (Postgraduate). Each student was required to rate ten Likert-scale questions on a 5-point scale. The first question was about the overall rating of the course, while the remaining nine questions evaluated various aspects of the course. We summarized these nine questions into topics (Table 1).

3.1 Comparison of Computer Science courses with other courses

We compare the mean scores for each course in the Computer Science department with those of other courses, both for on-campus and online students (Table 1). The "P-value" column represents the results of the Wilcoxon signed-rank test, which assesses the statistical significance of the differences between the mean scores.

Computer Science courses received lower mean scores for the course overall question compared to other courses, indicating that students were less satisfied with Computer Science courses in general. Computer Science courses also showed lower mean scores in several specific areas, including collaboration, communication, clarity, relevance, feedback, community, engagement, and quality, when compared to other courses.

Computer science courses have a unique curriculum design and delivery approach that may place greater emphasis on the learning of theoretical knowledge and technical skills while lacking some elements that emphasise collaboration, communication and interaction [4]. Such design differences may result in lower student satisfaction in collaboration, communication and other areas. Computer science courses typically involve more programming, and complex theoretical knowledge, and may require more independent learning and independent problem-solving skills [4]. This high level of difficulty and challenge may affect student satisfaction and lead to a perceived lack of clarity, relevance and quality of the course. Computer science courses may place greater emphasis on individual learning and independent exploration, with less emphasis on teacher-student interaction and feedback. This approach to teaching and learning may result in students feeling a lack of feedback, community and engagement.

3.2 Comparison between Computer Science courses

To analyze differences between the Computer Science courses, we compared the mean scores for each question across different stages (year level) within Computer Science. The results indicate several significant differences between courses (see Table 2).

Furthermore, we conducted an analysis to examine the impact of course type on student scores within the Computer Science courses. The analysis involved the categorization of 31 Computer

Table 1: Comparison of SET mean scores Between CS and other courses

Topic	Question	Computer Science		Other Courses		P-value
		Mean	Std.	Mean	Std.	(Wilcoxon signed-rank test)
Course Overall Question (inNZ)		3.759	0.546	4.168	0.658	**<0.001**
Course Overall Question (online)		3.771	0.548	4.264	0.841	**0.001**
Accessibility	Q1(inNZ)_It was easy to find the information and resources I needed on the Canvas course website	4.052	0.529	4.226	0.766	0.133
Collaboration	Q2(inNZ)_The learning environment provided me with opportunities to communicate and/or collaborate with my peers	3.634	0.498	4.214	0.733	**<0.001**
Communication	Q3(inNZ)_The learning environment allowed effective communication between teaching staff and students	3.898	0.506	4.242	0.607	**0.001**
Clarity	Q4(inNZ)_I was clearly informed how my learning would be assessed	4.018	0.511	4.252	0.661	**0.016**
Relevance	Q5(inNZ)_Assessments supported the aims of this course	4.110	0.435	4.337	0.620	**0.024**
Feedback	Q6(inNZ)_I received helpful feedback on my learning progress	3.664	0.575	4.039	0.700	**0.001**
Community	Q7(inNZ)_I felt part of a community of learners in this course	3.519	0.518	4.052	0.644	**<0.001**
Engagement	Q8(inNZ)_I felt I could stay motivated and engaged with my learning	3.599	0.560	3.955	0.711	**0.001**
Quality	Q9(inNZ)_I was satisfied with the quality of the small-group teaching (e.g. tutorial, laboratory, seminar, workshop) associated with this course	3.601	0.550	4.144	0.733	**<0.001**
Accessibility	Q1(online)_It was easy to find the information and resources I needed on the Canvas course website	4.146	0.608	4.290	0.544	0.181
Collaboration	Q2(online)_The learning environment provided me with opportunities to communicate and/or collaborate with my peers	3.828	0.843	3.914	0.647	0.236
Communication	Q3(online)_The learning environment allowed effective communication between teaching staff and students	3.838	0.780	4.103	0.744	**0.031**
Clarity	Q4(online)_I was clearly informed how my learning would be assessed	4.132	0.605	4.341	0.509	**0.024**
Relevance	Q5(online)_Assessments supported the aims of this course	4.048	0.462	4.391	0.624	**0.003**
Feedback	Q6(online)_I received helpful feedback on my learning progress	3.864	0.701	4.182	0.820	**0.043**
Community	Q7(online)_I felt part of a community of learners in this course	3.707	0.777	3.980	0.766	**0.047**
Engagement	Q8(online)_I felt I could stay motivated and engaged with my learning	3.619	0.696	4.031	0.987	**0.023**
Quality	Q9(online)_I was satisfied with the quality of the small-group teaching (e.g. tutorial, laboratory, seminar, workshop) associated with this course	3.725	0.706	4.115	0.903	**0.047**

Table 2: Comparison Between Different Stages of Computer Science Courses

Topic	Question	Stage 1		P-value	Stage 2		P-value	Stage 3		P-value	Stage 7		P-value
		Mean	Std.	(Wilcoxon)	Mean	Std.	(Wilcoxon)	Mean	Std.	(Wilcoxon)	Mean	Std.	(Wilcoxon)
Course Overall Question (online		4.092	0.336	0.364	3.306	0.408	**0.048**	4.242	0.355	**0.044**	3.386	1.157	0.243
Accessibility	Q1(inNZ)	4.577	0.255	**0.013**	4.217	0.420	0.822	3.857	0.251	**0.043**	3.815	1.160	0.751
Collaboration	Q2(inNZ)	3.370	0.621	0.270	3.488	0.397	0.316	3.604	0.440	0.285	3.906	1.018	**0.004**
Clarity	Q4(inNZ)	4.480	0.292	**0.014**	4.020	0.475	0.900	3.736	0.304	**0.018**	3.993	0.939	0.735
Relevance	Q5(inNZ)	4.520	0.290	**0.038**	4.088	0.581	0.726	4.128	0.280	0.514	3.862	0.841	0.434
Accessibility	Q1(online)	4.292	0.326	0.551	3.618	0.284	**0.012**	4.340	0.332	0.298	4.148	0.707	0.676
Collaboration	Q2(online)	3.788	0.277	0.979	3.242	0.266	**0.026**	4.130	0.362	0.103	3.873	0.893	0.912
Communication	Q3(online)	4.022	0.328	0.568	3.402	0.215	**0.030**	4.140	0.361	0.129	3.673	1.110	0.809
Clarity	Q4(online)	4.275	0.323	0.390	3.628	0.257	**0.011**	4.298	0.382	0.237	4.149	0.622	0.930
Relevance	Q5(online)	4.255	0.482	0.500	3.612	0.279	**0.022**	4.239	0.351	0.308	3.970	0.849	0.843
Feedback	Q6(online)	4.188	0.397	0.484	3.340	0.128	**0.021**	4.329	0.354	**0.042**	3.512	1.132	0.454
Community	Q7(online)	3.813	0.265	0.897	3.188	0.100	**0.018**	4.186	0.262	**0.014**	3.471	1.116	0.725
Engagement	Q8(online)	3.930	0.197	0.815	3.176	0.176	**0.032**	4.257	0.329	**0.006**	3.080	1.404	0.243
Quality	Q9(online)	4.062	0.345	0.518	3.128	0.239	**0.030**	4.228	0.329	0.066	3.329	1.314	0.524

Science courses into two distinct groups based on their class syllabus, namely Theory (n = 15) and Programming (n = 17). Subsequently, t-tests and Mann-Whitney U-tests were conducted to examine these course categories. The findings revealed that, among the online students, the mean scores for Course Overall Scores were significantly higher in theoretical courses (MeanT = 4.051) compared to programming courses (MeanP = 3.556), yielding a p-value of 0.025. Moreover, specific aspects of the courses were also examined, whereby it was observed that online theoretical courses exhibited significantly higher mean scores for Q3 (MeanT = 4.106, MeanP = 3.632, p=0.039 in t-test), Q6 (MeanT = 4.127, MeanP = 3.662, p=0.054 in Mann-Whitney U-test), Q7 (MeanT = 4.008, MeanP = 3.476, p=0.028 in t-test), Q8 (MeanT = 3.944, MeanP = 3.371, p=0.036 in Mann-Whitney U-test), and Q9 (MeanT = 3.971, MeanP = 3.525, p=0.051 in Mann-Whitney U-test) compared to programming courses.

In Stage 2 online course, overall scores and scores on all SET questions were significantly lower. This may be due to the wide span of difficulty from Stage 1 to Stage 2 and the fact that online teaching makes it difficult for students to access the appropriate resources and adapt, ultimately resulting in lower overall scores. In the Stage 3 online courses, the overall scores were significantly higher as well as the scores for feedback, community and engagement. This may be due to the fact that the smaller number of students in the courses allowed them to perform better in terms of the quality and interactivity of the teaching and learning, with students being more interested and engaged in the content and teaching methods. In face-to-face first-year courses (Phase 1), the mean scores for accessibility, clarity and relevance were significantly higher, which may be due to the fact that these courses performed well in terms of availability of teaching resources, clarity of presentation and relevance to students' interests.

3.3 Comparison between different Stage courses

We conducted a comprehensive analysis of SET scores of 272 courses from the science department, categorizing them based on stages and examining the differences in average scores across various evaluation questions (Table 3) There were no significant differences observed in the average scores of Stage 1 (typically first-year) courses compared to the other stages. In contrast, Stage 2 (typically second-year) courses displayed noticeable differences with consistently lower average scores in many aspects compared to the other stages. For Stage 3 (typically third-year) courses, a distinct pattern emerged between offline and online modes of instruction. In the offline context, Stage 3 courses demonstrated lower scores in the Accessibility aspect compared to the other stages, suggesting room for improvement in terms of course accessibility for students. However, in the online format, Stage 3 courses exhibited significantly higher scores in several aspects compared to the other stages. In the case of postgraduate courses (typically fourth-year), the analysis revealed that, excluding Accessibility and Clarity, these courses received higher average scores in all other evaluation dimensions. This indicates that postgraduate courses generally performed well in terms of student ratings. The higher scores could be attributed to the specialized nature of postgraduate education, which often involves more advanced and in-depth subject matter, as well as higher expectations from students.

It is unsurprising that postgraduate courses received higher average scores in all assessment dimensions except accessibility and clarity. This may be due to the fact that postgraduate education usually involves more advanced and in-depth subject content, and the smaller class sizes allow teachers to focus well on each student. In addition, postgraduate programmes often have a more specialised character and may attract an interested, more motivated and focused student body.

3.4 Impact of class size

We conducted a comparison of mean scores on different evaluation questions between large courses (student numbers >100) and small courses (student numbers <100) across the 272 courses (Table 4). For both the offline and online course overall question (in NZ), there was no significant difference in mean scores between large and small courses. This indicates that course size did not have a significant impact on students' perceptions of the overall course quality, both in the offline and online formats.

However, significant differences were observed in several areas. In the offline format, smaller courses scored higher on Collaboration, Communication, Relevance, Feedback, Community, Engagement, and Quality compared to larger courses ($p < 0.005$ for all). This indicates that students in smaller courses perceived higher levels of collaboration, communication, relevance, feedback, community, engagement, and quality compared to students in larger courses.

In the online format, smaller courses received higher scores in Collaboration (although not statistically significant at $p = 0.078$) and Communication (although not statistically significant at $p = 0.070$) compared to larger courses. However, smaller courses scored significantly higher on Relevance compared to larger courses ($p = 0.055$). No significant differences were found between large and small courses in the remaining aspects of the evaluation in the online format.

In smaller courses, interaction and collaboration between students may be more frequent, and may also provide teachers with more opportunities to focus on individual students and provide feedback. This close interaction and collaboration raises scores in collaboration and communication and also helps to improve students' perceptions of relevance and feedback. Smaller courses may encourage students to be more active participants in their learning. Students are more likely to voice their opinions, ask questions and participate in class discussions in a small class setting, which can increase their sense of engagement and motivation to learn, leading to higher scores in engagement, community and course quality.

4 DISCUSSION

The analysis of the SET rating data provided insights into the impact of different course features and disciplines on the scores of various components of student evaluation of teaching. Here are the key findings related to the research questions:

Table 3: Comparison between different Stage courses

Topic	Question	Stage 1		P-value	Stage 2		P-value	Stage 3		P-value	Stage 7		P-value
		Mean	Std.	(Wilcoxon)	Mean	Std.	(Wilcoxon)	Mean	Std.	(Wilcoxon)	Mean	Std.	(Wilcoxon)
Course Overall Question (inNZ		4.210	0.324	0.740	3.982	0.587	**0.036**	4.103	0.556	0.329	4.221	0.601	**0.042**
ourse Overall Question (online		4.155	0.516	0.366	4.071	0.636	0.187	4.358	0.492	0.052	4.187	0.628	0.817
Accessibility	Q1(inNZ)	4.297	0.300	0.790	4.243	0.481	0.841	4.148	0.519	**0.021**	4.239	0.622	0.124
Collaboration	Q2(inNZ)	4.131	0.436	0.318	3.933	0.494	**<0.001**	4.134	0.526	0.341	4.361	0.495	**<0.001**
Communication	Q3(inNZ)	4.224	0.338	0.385	4.043	0.488	**0.007**	4.177	0.494	0.316	4.335	0.584	**<0.001**
Clarity	Q4(inNZ)	4.320	0.276	0.566	4.163	0.479	0.295	4.168	0.567	0.233	4.274	0.557	0.199
Relevance	Q5(inNZ)	4.344	0.238	0.655	4.204	0.458	0.055	4.316	0.435	0.624	4.391	0.484	**0.024**
Feedback	Q6(inNZ)	3.948	0.384	0.085	3.758	0.619	**0.003**	3.989	0.517	0.458	4.187	0.646	**<0.001**
Community	Q7(inNZ)	3.932	0.396	0.190	3.820	0.523	**0.007**	3.933	0.476	0.090	4.230	0.581	**<0.001**
Engagement	Q8(inNZ)	3.939	0.345	0.665	3.699	0.602	**0.007**	3.919	0.519	0.610	4.067	0.632	**0.003**
Quality	Q9(inNZ)	4.119	0.413	0.874	3.897	0.614	**0.010**	4.076	0.485	0.183	4.221	0.642	**0.002**
Accessibility	Q1(online)	4.271	0.572	0.959	4.206	0.506	0.163	4.347	0.524	0.259	4.218	0.716	0.894
Collaboration	Q2(online)	3.904	0.701	0.698	3.669	0.773	**0.028**	4.112	0.699	**0.035**	3.833	0.991	0.695
Communication	Q3(online)	4.071	0.677	0.766	3.887	0.749	0.060	4.277	0.670	**0.027**	3.959	0.867	0.892
Clarity	Q4(online)	4.355	0.520	0.611	4.125	0.662	0.057	4.355	0.534	0.446	4.335	0.613	0.589
Relevance	Q5(online)	4.341	0.407	0.822	4.162	0.513	**0.037**	4.405	0.487	0.218	4.414	0.485	0.304
Feedback	Q6(online)	4.081	0.606	0.254	3.900	0.774	**0.031**	4.331	0.598	**0.027**	4.228	0.716	0.295
Community	Q7(online)	3.959	0.694	0.835	3.786	0.659	0.058	4.125	0.618	0.082	3.842	1.006	0.772
Engagement	Q8(online)	3.929	0.588	0.183	3.818	0.717	0.079	4.273	0.553	**0.001**	3.861	0.915	0.809
Quality	Q9(online)	4.007	0.651	0.295	4.017	0.709	0.640	4.211	0.697	0.074	4.003	0.755	0.761

Table 4: Comparison Between Different Sized Courses

Topic	Question	Large Courses (Students Size>100)		Small Courses (Students Size<100)		P-value (Wilcoxon test)
		Mean	Std.	Mean	Std.	
Course Overall Question (inNZ		4.074	0.446	4.145	0.629	0.082
ourse Overall Question (online		4.147	0.559	4.189	0.733	0.260
Accessibility	Q1(inNZ)	4.257	0.355	4.180	0.642	0.881
Collaboration	Q2(inNZ)	3.945	0.439	4.249	0.589	**<0.001**
Communication	Q3(inNZ)	4.095	0.384	4.256	0.582	**<0.001**
Clarity	Q4(inNZ)	4.222	0.344	4.227	0.610	0.156
Relevance	Q5(inNZ)	4.270	0.321	4.331	0.523	**0.021**
Feedback	Q6(inNZ)	3.850	0.420	4.069	0.665	**<0.001**
Community	Q7(inNZ)	3.769	0.416	4.102	0.590	**<0.001**
Engagement	Q8(inNZ)	3.807	0.458	3.967	0.640	**0.006**
Quality	Q9(inNZ)	3.949	0.479	4.148	0.642	**0.001**
Accessibility	Q1(online)	4.227	0.525	4.299	0.669	0.179
Collaboration	Q2(online)	3.822	0.745	3.981	0.866	0.078
Communication	Q3(online)	4.005	0.683	4.101	0.874	0.070
Clarity	Q4(online)	4.280	0.574	4.322	0.613	0.450
Relevance	Q5(online)	4.261	0.463	4.392	0.562	0.055
Feedback	Q6(online)	4.073	0.672	4.171	0.800	0.167
Community	Q7(online)	3.939	0.657	3.912	0.902	0.481
Engagement	Q8(online)	3.944	0.652	3.956	0.900	0.354
Quality	Q9(online)	4.007	0.685	4.077	0.842	0.252

4.1 RQ 1: Relationship between course features and student evaluations

Course Level (Stages): Stage 2 courses consistently received lower average scores compared to other stages across multiple aspects, suggesting that these courses may be more challenging and involve complex concepts. Students may require more time and effort to adapt to Stage 2 courses, resulting in lower ratings.

Course Type (Theory vs. Programming): Among online students in the Computer Science department, theoretical courses received significantly higher mean scores compared to programming courses in terms of Course Overall Scores and specific aspects like Communication, Feedback, Community, Engagement, and Quality. This indicates that students perceived higher satisfaction and quality in theoretical courses compared to programming courses in the online learning environment.

Course Size: In the offline format, smaller courses (student numbers <100) scored higher in Collaboration, Communication, Relevance, Feedback, Community, Engagement, and Quality compared to larger courses (student numbers >100). This suggests that students in smaller courses perceived higher levels of collaboration, communication, relevance, feedback, community, engagement, and quality compared to students in larger courses. In the online format, smaller courses also received higher scores in Collaboration, Communication, and Relevance, although the differences were not statistically significant.

4.2 RQ 2: Differences between Computer Science and other disciplines

Comparison with Other Courses: Computer Science courses received lower mean scores for the course overall question compared to other courses, indicating lower student satisfaction. Computer Science courses also scored lower in several areas, including Collaboration, Communication, Clarity, Relevance, Feedback, Community, Engagement, and Quality, when compared to other courses. Both on-campus and online students showed lower mean scores for Computer Science courses compared to other courses in several areas.

Course Type (Theory vs. Programming): Among online learning students in the Computer Science department, theoretical courses received significantly higher mean scores in Course Overall Scores, as well as in specific aspects such as Communication, Feedback, Community, Engagement, and Quality, compared to programming courses. This suggests that students perceived higher satisfaction and quality in theoretical courses compared to programming courses in the online learning environment.

Course Level (Stages): Within the Computer Science department, Stage 2 Computer Science courses received significantly lower mean scores for Online Overall Scores and all SET questions compared to other stages. In contrast, Stage 3 Computer Science courses (both online and offline) received significantly higher mean scores for Online Overall Scores, Feedback, Community, and Engagement compared to other stages. The differences in mean scores indicate variations in student satisfaction and perceived quality across different stages of Computer Science courses.

5 THREATS TO VALIDITY

While every effort was made to conduct a rigorous analysis, several possible threats to the validity of the findings should be acknowledged and taken into account when interpreting the results of the study. The SET rating data were collected from courses in a single University. This limited scope may not be fully representative of the diversity of programs and students at different institutions or in different semesters. Caution should be exercised in generalizing the findings beyond the specific scope of this study. Also, the SET assessment process is voluntary, and not all students participated in the assessment. It is possible that students who had a positive or negative experience with the course were more willing to provide feedback, leading to possible bias in the data collected.SET ratings were collected at the end of the course, and student experiences and perceptions may have been influenced by the timing of the course, the assessment, or interactions with the instructor during the assessment. These timing factors may lead to biased feedback.

6 CONCLUSIONS

This research provides valuable insights into the complex relationship between course characteristics, student evaluation of teaching measures, and disciplinary differences in higher education, with a focus on Computer Science courses. Significant factors influencing SET scores were identified, including course level, type, and size. Stage 2 courses received lower ratings, possibly due to their increased complexity. Theoretical courses, especially in online formats, were associated with higher satisfaction compared to programming courses. Smaller course sizes correlated with higher scores across various aspects, highlighting the importance of class size in student perceptions.

Computer Science courses received lower overall ratings and scores in key areas compared to other disciplines. Both on-campus and online students showed lower mean scores for Computer Science courses, suggesting the need for improvements in teaching and learning practices. The study indicates the importance of tailoring teaching strategies to optimize student experiences and outcomes, particularly in Computer Science courses. Limitations include the single-institution focus and analysis from a single semester, limiting generalizability. Longitudinal studies could provide further insights into trends over time.

In conclusion, this research contributes to the understanding of the relationship between student evaluation of teaching, course characteristics, and disciplinary differences. It emphasizes the need for continuous assessment and adaptation of teaching practices to create positive learning environments. Tailoring approaches to specific course needs can enhance the quality of education and improve student experiences. Future research should explore additional factors influencing student evaluation of teaching scores and consider longitudinal data for a more nuanced understanding of higher education dynamics.

REFERENCES

[1] Henry A. Hornstein. 2017. Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education* 4, 1 (Jan. 2017), 1304016. https://doi.org/10.1080/2331186X.2017.1304016 Publisher: Cogent OA _eprint: https://www.tandfonline.com/doi/pdf/10.1080/2331186X.2017.1304016.

[2] Jim Ivins, Brian R Von Konsky, David Cooper, and Michael Robey. 2006. Software Engineers and Engineering: A Survey of Undergraduate Preconceptions. In *Proceedings. Frontiers in Education. 36th Annual Conference*. 6–11. https://doi.org/10.1109/FIE.2006.322364 ISSN: 2377-634X.

[3] Antti Knutas, Timo Hynninen, and Maija Hujala. 2021. To Get Good Student Ratings should you only Teach Programming Courses? Investigation and Implications of Student Evaluations of Teaching in a Software Engineering Context. In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*. 253–260. https://doi.org/10.1109/ICSE-SEET52601.2021.00035

[4] Michael Morgan, Jane Sinclair, Matthew Butler, Neena Thota, Janet Fraser, Gerry Cross, and Jana Jackova. 2018. Understanding International Benchmarks on Student Engagement: Awareness and Research Alignment from a Computer Science Perspective. In *Proceedings of the 2017 ITiCSE Conference on Working Group Reports (ITiCSE-WGR '17)*. Association for Computing Machinery, New York, NY, USA, 1–24. https://doi.org/10.1145/3174781.3174782

[5] Wolfgang Stroebe. 2020. Student Evaluations of Teaching Encourages Poor Teaching and Contributes to Grade Inflation: A Theoretical and Empirical Analysis. *Basic and Applied Social Psychology* 42, 4 (July 2020), 276–294. https://doi.org/10.1080/01973533.2020.1756817 Publisher: Routledge _eprint: https://doi.org/10.1080/01973533.2020.1756817.

[6] Bob Uttl, Carmela A. White, and Daniela Wong Gonzalez. 2017. Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation* 54 (Sept. 2017), 22–42. https://doi.org/10.1016/j.stueduc.2016.08.007