# Can Early Programming Performance Predict Computer Science Students' Success?

**Debby Erce Sondakh[1],** Stenly Richard Pungus[2]
[1,2]Faculty of Computer Science, Universitas Klabat
e-mail: [1]**debby.sondakh@unklab.ac.id**, [2]stenly.pungus@unklab.ac.id

***Abstract***

*Investigating the possibility of lower-level computer programming courses predicting future performance of computer science students has received a lot of attention from scholars. This study mainly aimed to predict the success of computer science students based on their performance in the first two computer programming courses, namely Computer Programming I and Computer Programming II. The study employed a quantitative correlational design. Six years of data from graduating students were analyzed. The results demonstrate that the better the grade on Computer Programming I and II, the shorter the study duration. When further analysis was conducted to find out whether gender diversity exists, the results demonstrated that in Computer Programming I and II, female students outperformed males. Statistically, this difference was only significant in Computer Programming I. A greater proportion of female students graduated on time, yet it is not statistically significant.*

***Keywords***—Computer Programming, Predict Students' Success, Study Duration, Gender Gap

***Abstrak***

*Penelitian tentang apakah mata kuliah pemrograman komputer dapat memprediksi kinerja masa depan mahasiswa ilmu komputer telah menarik minat yang cukup besar dari kalangan akademisi. Tujuan utama penelitian ini adalah untuk memprediksi keberhasilan mahasiswa ilmu komputer berdasarkan kinerja (nilai) mereka dalam dua mata kuliah pemrograman komputer tahun pertama, yaitu Pemrograman Komputer I dan II. Penelitian ini menggunakan desain kuantitatif korelasional. Data yang dianalisis adalah data lulusan 6 tahun. Hasil penelitian menunjukkan bahwa semakin baik nilai Pemrograman Komputer I dan II, maka durasi studi semakin pendek. Analisis lebih lanjut dilakukan untuk mengetahui apakah terdapat perbedaan berdasarkan jenis kelamin terhadap nilai kedua mata kuliah dan durasi kuliah, hasilnya menunjukkan bahwa dalam Pemrograman Komputer I dan II, mahasiswa perempuan lebih unggul dibanding laki-laki. Secara statistik, perbedaan ini hanya signifikan pada Pemrograman Komputer I. Proporsi mahasiswa perempuan yang lulus tepat waktu lebih tinggi, namun tidak signifikan secara statistik.*

***Kata Kunci***—Computer Programming, Predict Students' Success, Study Duration, Gender Gap

## 1. INTRODUCTION

As computing technology continues to shrink in size while increasing in speed and processing capacity while also continuing to become more reasonably priced, it will soon be possible to apply it to nearly every element of human endeavor. The following examples reflect a world where computing technology is everywhere: In the business field, it is used to keep an up-to-date record of transactions and generate reports for use in research analysis and marketing efforts. In the medical industry, it is used to streamline and automate the administrative processes, health computing technology has revolutionized many areas, including hospital information systems, medical data analysis and imaging, the care of critically ill patients, and computer-assisted therapy. In agriculture, including image analysis to deal with biological variability between items of produce and hardware development to control field operations in produce grading, crop yield prediction, and crop and animal management In education, computing technology is pushing the boundaries of knowledge in physics, chemistry, biology, and many other majors. In entertainment, it is used to make movies, video games, and other media accessible on a wide variety, and so forth. These are just a few example of who computers are integrated into our daily live. Accordingly, it shouldn't be shocking to hear the claim that "one may major in computer science and do anything" [1]. Thus, students who want to keep up with the constantly evolving workforce may consider taking a computer science course, as this discipline is a prominent innovation engine.

The subfield of computing known as computer science places an emphasis on coding, mathematical procedures, and organizational structures of data. Abstraction, complexity, and evolutionary change are recurring topics in computer science that are strongly connected to a person's ability to program and develop software [2]. The development of abilities and fluency in computer programming is an important component of the computer science major. This refers to the ability to create and implement sets of instructions that allow a computer to carry out a certain task or find solutions to problems [3].

The programming courses in the computer science curriculum are structured as a series. The beginning programming courses are requirements for upper-level courses such as algorithms and complexity, software development principles, and software engineering, because the knowledge gained in prerequisite courses at a lower level serves as the foundation for courses at a higher level. There has been a lot of discussion in research community over whether or not introductory programming courses can accurately predict students' success in computer science. The purpose of the research presented in [4] was to determine whether or not computer science students would be successful in a software engineering course by analyzing their performance in earlier levels of the curriculum, specifically Computer Science I and Computer Science II, as well as Data Structures and Object-oriented Problem Solving. It is concluded that the four predictors account for the variance in students' performance in the Software Engineering course, and it was discovered that Computer Science II has a strong and positive causal link with students' success in the Software Engineering course.

Comparatively, a previous study indicated that students who are successful in the Software Engineering course earned grades higher than a C in the first two programming courses, which cover the fundamentals of programming (variables, loops, procedures, arrays, and object oriented programming ideas [5]. Similarly, the cumulative GPA of graduating students was predicted based on their performance in programming-related courses during their first three semesters (Computer Science I, Computer Science II, and Data Structure) [6]. The study suggests that Computer Science I was a good predictor of a student's success in completing a Computer Science degree, namely Computer Science I. With a higher grade in Computer Science I, the student will graduate with a higher cumulative GPA.

In addition, the researchers are eager to determine whether there are gender variations between male and female students in programming courses and duration of study. Given the chronic gender disparity in computer science education, it is always fascinating to examine gender diversity in this profession. Computer science education continues to struggle with gender diversity [7]. Men and women may both have a positive attitude toward programming, but males have more [8]–[10]. This may be related to women's lower level of interest and self-efficacy in computing [11] that men perform better than women [12] and the need for a gender-specific strategy to teaching computer science related concepts arise [13]. Other studies, on the other hand, found no significant differences between male and female programming learners or programming concepts [14]–[16].

The purpose of our research was to investigate the success of Computer Science majors, on the topic of study duration, based on their performance (grades) in early programming classes, particularly in a small Computer Science department at an Indonesian university. This study also looking into the issue of gender inequalities in computer programming course and study duration. By employing academic data from Computer Science students, the following research questions were examined:
Question 1: What early programming courses best predict students' study duration?
Question 2: To what extent is programming course performance and study duration different for female and male students?

## 2. METHOD

In this investigation, a quantitative correlational methodology was employed. A correlational design was utilized owing to its capacity to examine the effects of two or more variables. It thus allows for the prediction of students' study duration based on their performance in programming courses, which is the purpose of this study.

### *2.1 Data Collection*

For the purpose of this study, the researcher gathered the academic data of all Computer Science students from the university's academic information system, for all semesters from 2013 to 2018 academic years. In those years, 655 enrollments have taken place in the Computer Science department. In this study, only the grades of students who already had graduated were included. These students were eliminated because they did not complete their degrees. In addition, Pass, Incomplete, Withdraw Pass, and Withdraw Failed grades were not considered in this analysis. Hence, 368 students were included in the analysis: 249 male and 119 female. In detail, the data collected include: demographic (gender), academic data: grade of Computer Programming I and Computer Programming II courses, and outcome data: number of semester taken to graduate. This study employes the following numeric grade scale: A as 4.0, A- as 3.7, B+ as 3.3, B as 3, B- as 2.7, C+ as 2.3, C as 2, D as 1, and F as 0.

## 3. RESULTS AND DISCUSSIONS

Table 1 displays the proportion of students who received an A, A-, B+, B-, C+, C-, D, or F in Computer Programming I and Computer Programming II. The majority of students received an A- in both subjects, with corresponding percentages of 25.5% and 20%. On average, students obtained a lesser mark in Computer Programming II: the proportion of students who achieved an A, A-, or B+ in Computer Programming I declined in the succeeding course..

Table 1. Percentage of students that obtained each grade in the first two programming classes

| | Computer Programming I | Computer Programming II |
|---|---|---|
| A | 21.7 | 16.8 |
| A- | 25.5 | 20.4 |
| B+ | 16.3 | 12.5 |
| B | 12.5 | 15.5 |
| B- | 9.8 | 11.4 |
| C+ | 9.8 | 13.9 |
| C | 3 | 6 |
| C- | 1.4 | 3.5 |
| D | 0 | 0 |
| F | 0 | 0 |

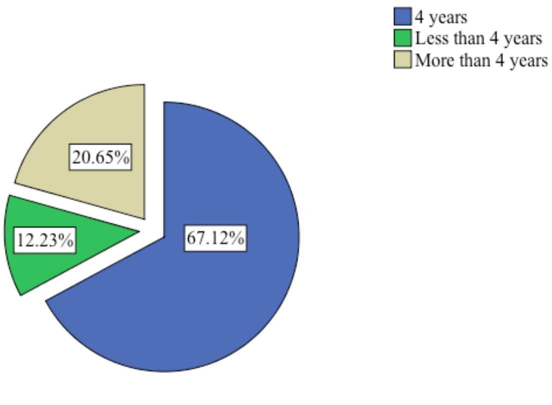


Figure 1. Study Duration Percentage

The program is designed to allow students to complete a degree in computer science in four years or 8 semesters. Initial investigation indicates that the average time of study for the 368 graduates was 8.3 semesters. Of them, 67.12% earned their degree in four years, 12.23% did it in less than four years, and 20.65% took longer than four years (see Figure 1). Details are provided in Table 2. In total, 302 students successfully completed their degrees on time.

Table 2. Semesters taken to graduate

| Semester | Frequency | Percentage |
|---|---|---|
| 6 | 1 | 0.3 |
| 7 | 44 | 12 |
| 8 | 247 | 67.1 |
| 9 | 38 | 10.3 |
| 10 | 23 | 6.3 |
| 11 | 6 | 1.6 |
| 12 | 8 | 2.2 |
| 13 | 1 | 0.3 |

Research question 1: What early programming courses best predict students' study duration?

To begin answering this issue, a One-way Analysis of Variances (ANOVA) was undertaken to compare the averages of the total number of semesters required to successfully complete the course (study length) with the grade on Computer Programming I as the fixed factor. The result indicates a statistically significant main effect: [$F(7,360) = 7.096$, p 0.05]. Table 3 presents the Computer Programming I grades and mean study duration. Similarly, significant main effect of study duration with grade on Computer Programming II was found at [$F(7,360) = 12.264$, $p < 0.05$].

Table 3. Grade on Computer Programming I versus study duration

| Grade | Study Duration | |
|---|---|---|
| | Mean | Std.Dev |
| A | 7.88 | 0.769 |
| A- | 8.06 | 0.700 |
| B+ | 8.15 | 0.917 |
| B | 8.59 | 1.127 |
| B- | 8.42 | 0.996 |

| | | |
|---|---|---|
| C+ | 9.00 | 1.394 |
| C | 8.64 | 1.120 |
| C- | 8.60 | 0.894 |

Table 4. Grade on Computer Programming II versus study duration

| Grade | Study Duration | |
|---|---|---|
| | Mean | Std.Dev |
| A | 7.76 | 0.645 |
| A- | 8.02 | 0.615 |
| B+ | 8.02 | 0.907 |
| B | 8.30 | 0.963 |
| B- | 8.36 | 0.805 |
| C+ | 8.57 | 0.965 |
| C | 8.91 | 1.411 |
| C- | 9.85 | 1.625 |

Similarly, significant main effect of study duration with grade on Computer Programming II was found at [$F(7,360) = 12.264$, $p < 0.05$] (see Table 4). It is notable that the better the grade on Computer Programming II, the lesser the mean of study duration. In contrast, this is not pertain to Computer Programming I. Students with lesser grades, such as B- and B, graduated earlier than those with better grades. Similar patterns were seen among students with grades of C/C- and C+. In addition, Table 3 and Table 4 reveal that a minimum B+ grade for both courses is required for students to graduate on time. Those who received less than a B+ took more than four years to graduate on average.

Next, the researchers examine the correlation between graduates' achievement in the first two programming courses (Computer Programming I and Computer Programming II) and their study time.

Table 5. Linearity test on Computer Programming I and Computer Programming II grades against Study Duration

| | | | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Study_Duration1 * Computer Programming I | Between Groups | (Combined) | 7 | 2.426 | 8.611 | 0.000 |
| | | Linearity | 1 | 14.397 | 51.11 | 0.000 |
| | | Deviation from Linearity | 6 | 0.43 | 1.527 | 0.168 |
| | | Within Groups | 360 | 0.282 | | |
| | | Total | 367 | | | |
| Study_Duration1 * Computer Programming II | Between Groups | (Combined) | 7 | 3.211 | 12.053 | 0.000 |
| | | Linearity | 1 | 20.881 | 78.377 | 0.000 |
| | | Deviation from Linearity | 6 | 0.266 | 0.999 | 0.426 |
| | | Within Groups | 360 | 0.266 | | |
| | | Total | 367 | | | |

Correlation analysis is useful for determining the existence of connections. Prior to doing correlation analysis, a linear relationship was examined between each Computer Programming I and Computer Programming I grade and study duration. Table 5 shows, linear relationship was found when analyzing, respectively, Computer Programming I and Computer Programming II against that on Study Duration: [$F(6, 360) = 1.527$, $p > 0.05$] and [$F(6, 360) = 0.999$, $p > 0.05$].

Table 6. The Association Between Performance in Programming and Study Duration

| | | Computer Programming I | Computer Programming II | Study Duration |
|---|---|---|---|---|
| Computer Programming I | Pearson Correlation | 1 | .440** | -.349** |
| | Sig. (2-tailed) | | .000 | .000 |
| | N | 368 | 368 | 368 |
| Computer Programming II | Pearson Correlation | .440** | 1 | -.420** |
| | Sig. (2-tailed) | .000 | | .000 |
| | N | 368 | 368 | 368 |
| Study Duration | Pearson Correlation | -.349** | -.420** | 1 |
| | Sig. (2-tailed) | .000 | .000 | |
| | N | 368 | 368 | 368 |

A Pearson product-moment correlation was then carried out to discover the association between Computer Programming I and Computer Programming I grades and study time. The association between each programming course and duration of study is negative and statistically significant, as seen in Table 6:
Computer Programming I against Study Duration: $r = -0.349$, $n = 368$, $p = 0.000$
Computer Programming II against Study Duration: $r = -0.420$, $n = 368$, $p = 0.000$
As the grade on each programming course increase, the study duration decreases. These results are also corroborated in Table 3 and Table 4. However, the coefficient correlation *r* value of both indicate weak association.

Further analysis was then conducted to identify programming courses that would predict study duration. Regression analysis provides insight into the impact that programming course grades have on students' study duration. The analysis results are shown in Table 7. The model's multiple correlation coefficient *R* comes out at a value of 0.458, which implies that 21% of the variance in the dependent variable Study Duration was accounted for by the combination of the two independent variables Computer Programming I and Computer Programming II. The *F*-ratio test result indicates that the overall regression model is a good fit for the data [$F(2, 365) = 48.42$, $p$  0.05]. When further analysis was conducted on each individual course grade, the statistical significance of Computer Programming I and Computer Programming II grades indicated that both Computer Programming I ($t = -3.922$, $p$  0.05) and Computer Programming II ($t = -6.378$, $p$  0.05) added statistically significantly to the prediction study duration. The unstandardized coefficient *B* of Computer Programming I (-0.188) indicates that for each Computer Programming I grade increase, there is a decrease in study duration of 0.188 semesters. Likewise, the unstandardized coefficient *B* of Computer Programming II (-0.275) indicates that for each Computer Programming II grade increase, there is a decrease in study duration of 0.275 semesters.

Table 7 Regression Analysis Predicting Computer Science Students' Success

| Item | Value | |
|---|---|---|
| Regression factor (R) | 0.458 | |
| R Square ($R^2$) | 0.210 | |
| Adjusted R Square | 0.205 | |
| F-test | 48.42 | |
| Grade / Study Duration | Computer Programming I | Computer Programming II |

| | | |
|---|---|---|
| Coefficient | -0.188 | -0.275 |
| *t*-value | -3.922 | -6.378 |
| *p*-value | 0.000 | 0.000 |

Research question 2: To what extent is programming course performance and study duration different for female and male students?

Comparing means is one method for responding the question. Thus, an independent-sample t-test was undertaken to compare the means of female and male students' programming grades and study duration. The graphs in
Figure 2 and Figure 3 illustrates mean comparisons of female and male students on Computer Programming I and Computer Programming II grades and study duration, respectively.

Students' study duration was shown to have an inverse relationship with their performance in both Computer Programming I and II. Better performance in computer programming courses, and especially in computer programming II, correlates with less time spent learning. Since these two courses set the groundwork for other succeeding programming courses in the curriculum, such as data structure, object-oriented programming, and so on, it implies that students must comprehend the material and receive a high grade in order to graduate on time. This finding expands upon prior research on the use of first-year programming course grades as a predictor of performance in upper-level CS courses [17] like Software Engineering [5] or GPA upon graduation [4], using first programming course grades. In short, performance in computer programming courses serve as predictor to students' success in term of study duration.

Looking deeper into the data, a minimum of B+ (3.3) grade is required to graduate on time. Given the regression coefficient value and the model's goodness-of-fit measures, which are moderately low, Computer Programming I and II grades can be usd as 'early warning' to help students avoid not being able to graduate on time, however, they cannot be used as the basis to recommend that student reconsider their choice of major. Further analysis is required to assess the association between low grades in computer programming and students' withdrawal or failure rate.

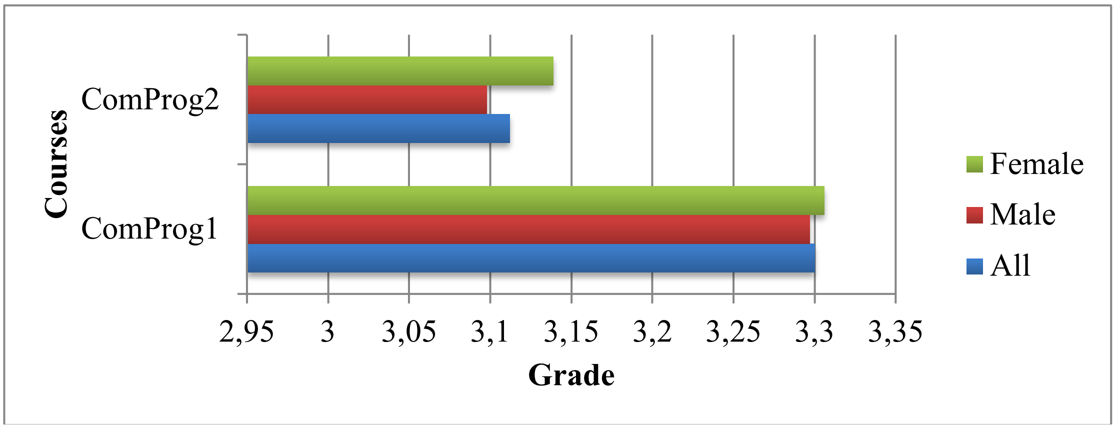


Figure 2. Comparing Female and Male Students' Computer Programming I and Computer Programming II Grades

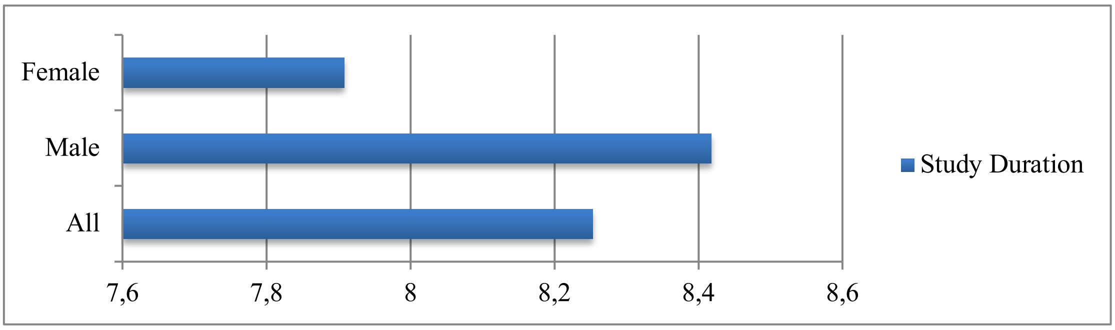


Figure 3. Comparing Female and Male Students' Study Duration

As can be seen in Figure 2, the means of female students in both Computer Programming I (3.306) and Computer Programming II (3.139) are higher than the means of male students (Computer Programming I 3.297 and Computer Programming II 3.098). When comparing the means of female students in these two courses, greater values are found than when comparing the means of all students (Computer Programming I 3.300 and Computer Programming II 3.112). As for male students, their mean Computer Programming I and Computer Programming II grades are below the mean for all students. Figure 3 depicts a similar trend, in which the mean study time of female students (7.908) is less than that of male students (8.253) and all students (8.253). On the other hand, the mean study duration of male students (8.418) is longer than that of all students.

When the means of programming course grades are further examined significant differences is found in only Computer Programming II grade [$t(236.8) = 0.567$, $p < 0.05$ ), as shown in Table 8. In the same way, Table 9 shows statistically significant difference is found in study duration, in which on average female students graduated ealier than male students [$t(365.4) = -5.892$, $p < 0.00$].

Table 8. Programming course grades differences

| | Gender | M | SD | t | df | Sig.(2-tailed) |
|---|---|---|---|---|---|---|
| Computer Programming I | Female | 3.306 | 0.552 | 0.133 | 366 | 0.895 |
| | Male | 3.297 | 1.108 | | | |
| Computer Programming II | Female | 3.139 | 0.620 | 0.567 | 236.8 | 0.000 |
| | Male | 3.098 | 0.712 | | | |

Table 9. Study duration differences

| | Gender | M | SD | t | df | Sig.(2-tailed) |
|---|---|---|---|---|---|---|
| Study Duration | Female | 7.908 | 0.552 | -5.892 | 365.4 | 0.000 |
| | Male | 8.418 | 1.108 | | | |

Surprisingly, this findings revealed that the mean grades of female students in Computer Programming I and II are higher than those of male students. There were statistically significant differences in Computer Programming I. While it is true that there was no statistically significant difference between the grades of female and male students in Computer Programming II, such outcomes could indicate a constriction of gender diversity in computer science, specifically programming, as well as the same potential and abilities that both men and women possessed in understanding programming concepts, as previous research [15]–[17] concluded.

In terms of study duration, it was shown that female students outperformed their male counterparts. On average, female students graduate in fewer than four years. This may suggest that female students are more capable academically than male students. However, other variables, such as the number of summer schools attended by the students, may impact the amount of time required to complete a degree. These variables are not included in the analysis. At the university where this research is being undertaken, summer school is scheduled for students who earned a grade of C- or lower in any course and need to retake it. Some students take advantage of this opportunity to enroll in classes they should have taken in the next normal semester, therefore shortening the duration of their studies.

## 4. CONCLUSION

Through this study, the researchers have attempted to predict the success (graduation on time) of computer science department students by analyzing their first two programming courses, i.e. Computer Programming I and II. The following are highlighted the findings:

- Both courses have a statistically significant effect on the duration of study, the better the grade in Computer Programming I and II, the shorther the study duration.
- In Computer Programming I and II, female students did better than male students. However, this difference was only statistically significant in Computer Programming I.
- Proportion of female students who graduated on time is higher.

These findings have implication for computer science education, as a general, and in particular programming teaching and learning. Being able to predict students' study duration since their first year enables a computer science department to recommend specific study plan, especially for students with lower Computer Programming grades. Accordingly, the students might as well be benefited from the pattern resulting from the analysis, wherein they can do self-assessment in response to their own performance in these two programming classes they have to take in the first year. It might help in making decision for planning and managing their study duration to complete a computer science major degree.

## 5. FUTURE WORK

In addition to the optimistic findings of this study, several recommendations for further research have been compiled. The first is to incorporate more programming-related courses in order to predict the students' study duration. As mentioned earlier, programming courses within the computer science curriculum are organized in a series. Computer programming is only an introductory programming course that lays the groundwork for more advanced courses. By examining all programming courses collectively, a more accurate graduation prediction model may be created. Secondly, to predict students' academic success holistically, other variables such as number of summer schools attended, attitudes and other psychological factors may also be considered. Finally, further research can be conducted to come up with an effective pedagogy for teaching computer programming, especially to students in developing countries such as Indonesia where technological infrastructure is lacking. This will help students comprehend the essentials of computer programming, learn them better, and employ them in more advanced courses.

## REFERENCES

[1] J. M. Wing, 2006, Computational thinking, *Commun. ACM*, vol. 49, no. 3, pp. 33–35.

[2] ACM and IEEE-CS, 2021, Computing Curricula 2020 paradigms for Global Computing Education, New York.

[3] R. Scherer, F. Siddiq, and B. Sánchez Viveros, 2020, A meta-analysis of teaching and learning computer programming: Effective instructional approaches and conditions, *Comput. Human Behav.*, vol. 109, no. 0318.

[4] S. M. Khan, M. Ibrahim, and S. A. Haider, 2021, Performance Prediction of Computer Science Students in Capstone Software Engineering Course Through Educational Data Mining, *ASEE Annu. Conf. Expo. Conf. Proc.*, no. 479.

[5] D. A. Trytten and A. McGovern, 2017, Moving from managing enrollment to predicting student success, *Proceedings - Frontiers in Education Conference, FIE*, 2017, vol. 2017-Octob, pp. 1–9.

[6] A. N. Kumar, 2018, Predicting student success in computer science - A reproducibility study, *Proceedings - Frontiers in Education Conference, FIE*, 2019, vol. 2018-Octob, pp. 1–6.

[7] L. J. Sax *et al.*, 2017, Anatomy of an enduring gender gap: the evolution of women's participation in Computer Science, *J. Higher Educ.*, vol. 88, no. 2, pp. 258–293.

[8] M. Baser, "Attitude, gender and achievement in computer programming," *Middle East J. Sci. Res.*, vol. 14, no. 2, pp. 248–255, 2013.

[9] H. Suwono, R. Fachrunnisa, C. Yuenyong, and L. Hapsari, 2019, Indonesian students' attitude and interest in STEM: An outlook on the gender stereotypes in the STEM field, *J. Phys. Conf. Ser.*, vol. 1340, no. 1.

[10] Z. Cai, X. Fan, and J. Du, 2017, Gender and attitudes toward technology use: A meta-analysis, *Comput. Educ.*, vol. 105, no. 2017, pp. 1–13.

[11] S. Beyer, 2014, Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades, *Comput. Sci. Educ.*, vol. 24, no. 2–3, pp. 153–192.

[12] Z. Gao, Y. Zhang, R. Zhang, X. Sun, and J. Feng, 2022, Do Gender or Major Influence the Performance in Programming Learning? Teaching Mode Decision Based on Exercise Series Analysis, *Comput. Intell. Neurosci.*, vol. 2022.

[13] C. Angeli and N. Valanides, 2020, Developing young children's computational thinking with educational robotics: An interaction effect between gender and scaffolding strategy, *Comput. Human Behav.*, vol. 105.

[14] M. S. Gunbatar and H. Karalar, 2018, Gender differences in middle school students' attitudes and self-efficacy perseptions towards mBlock programming, *Eur. J. Educ. Res.*, vol. 7, no. 4, pp. 925–933.

[15] S. Atmatzidou and S. Demetriadis, 2016, Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences, *Rob. Auton. Syst.*, vol. 75, pp. 661–670.

[16] D. E. Sondakh, S. R. Pungus, and E. Y. Putra, 2022, Indonesian Undergraduate Students' Perception of Their Computational Thinking Ability, *Cogito Smart J.*, vol. 8, no. 1, pp. 68–80.

[17] S. Malla, J. Wang, W. Hendrix, and K. Christensen, 2019, Predicting Success for Computer Science Students in CS2 using Grades in Previous Courses, *Proc. - Front. Educ. Conf. FIE*, vol. 2019-Octob.