Predicting CGPA Using Multiple Linear Regression

M. Abdullah Ali – 23i-2523

Muddassir Asghar – 23i-2577

FAST NUCES Islamabad – Data Science – Section A

MT-2005 : Probability and Statistics

Author Note

M. Abdullah Ali and Muddassir Asghar are undergraduate students at FAST NUCES Islamabad. This report was created as part of the requirements of the semester project for the course MT-2005.

Correspondence regarding this research should be addressed to:

M. Abdullah Ali
Email: i232523@isb.nu.edu.pk

Muddassir Asghar
Email: i232577@isb.nu.edu.pk

Abstract

This research explores the prediction of students' Cumulative Grade Point Average (CGPA) using a Multiple Linear Regression Model (MLRM). Primary data were collected through in-person surveys and online forms, examining variables including semester level, weekly study hours, attendance percentage, and daily screen time. The analysis revealed that attendance is the most significant positive predictor of CGPA, while screen time negatively impacts academic performance. The model accounted for approximately 47% of the variance in CGPA, with three predictors showing statistical significance ($p < 0.05$) and semester marginally falling outside this threshold ($p = 0.085$). These findings emphasize the critical role of class attendance and efficient study habits in academic success, offering practical insights for students and educators. The report discusses limitations such as unexplained variance and variable scope, proposing directions for future research to refine predictive accuracy and incorporate additional factors.

Keywords: CGPA Prediction, Multiple Linear Regression, Academic Performance, Primary Data, Educational Research

Predicting CGPA Using Multiple Linear Regression

1. Introduction

Academic performance is a critical measure of students' progress and success in educational settings. The Cumulative Grade Point Average (CGPA) serves as a universal indicator of this performance. Predicting CGPA can provide valuable insights for educators and students, helping identify key factors influencing academic success. This project aims to construct a Multiple Linear Regression Model (MLRM) to predict CGPA using factors such as semester, weekly study hours, attendance, and daily screen time. By analyzing these predictors, the study seeks to provide actionable recommendations for improving academic outcomes.

2. Literature Review

The prediction of academic performance, often measured by Cumulative Grade Point Average (CGPA), has warranted significant attention in educational research. Multiple studies have explored diverse factors influencing CGPA, including behavioral habits, academic engagement, and technological influence.

Attendance and Academic Performance

Attendance consistently emerges as a critical factor in shaping CGPA. Studies at Gaziosmanapaşa University and Kolej Universiti Insaniah highlighted the positive association between attendance and academic outcomes (Erdem et al., 2007). Students who attend classes regularly benefit from increased engagement, direct access to instructional content, and active participation in academic discussions, all of which are conducive to better performance. These findings establish attendance as an essential predictor, aligning with its inclusion in this study.

Study Habits and Academic Performance

Study habits, particularly the number of hours dedicated to studying, have long been considered central to academic success. Research from Mawlana Bashani Science and Technology University emphasized that the quality of study time, rather than the quantity, may play a more significant role in determining outcomes (Akhter & Rahaman, 2020). The study found that while study hours generally have a positive impact, excessive study time may reflect inefficiencies or counterproductive stress. This nuanced understanding of study habits is critical for interpreting their influence on CGPA.

Screen Time and Digital Distractions

Technological advancements and increased digital access have introduced a new dimension to academic research. A study by Mahmud et al. (2023) revealed that prolonged online gaming, particularly playing for more than 30 hours per week, negatively impacts CGPA. The study also highlighted associated issues such as reduced study hours and physical activity, as well as skipping classes. These findings align with broader research indicating that excessive screen time diminishes academic performance by fostering distractions and undermining self-control. The adverse effects of screen time underscore its inclusion as a predictor in this research.

Semester-Level Challenges

***T**he progression through academic semesters introduces additional challenges that can influence CGPA. Advanced coursework, higher expectations, and external pressures such as internships and job preparation contribute to increased academic demands in later semesters. While this area is less frequently studied directly, research from Mawlana Bhashani Science and Technology University indirectly suggests that academic rigor intensifies with progression, potentially leading to lower performance metrics (Akhter & Rahaman, 2020).

Synthesis and Research Gap

The reviewed literature collectively highlights key predictors of CGPA, including attendance, study habits, screen time, and semester level. While previous studies provide valuable insights, gaps remain in understanding the interplay of these factors within a unified predictive framework. This research aims to address this gap by using a Multiple Linear Regression Model to evaluate the combined influence of these variables, providing actionable insights.

3. Regression Model and Variables

The study utilized a Multiple Linear Regression Model (MLRM) to predict CGPA. The dependent variable, CGPA, represents a student’s cumulative academic performance. Four independent variables were included:

1. ***Semester:***

   Semester represents the student's current semester. This ordinal variable captures the number of challenges faced as students advance through their degree.

2. ***Study Hours (hours/week):***

   Study hours measure the average weekly time spent studying. While higher study hours are generally expected to positively impact performance, diminishing returns of inefficiency in study habits may lead to unexpected trends.

3. ***Attendance (%):***

   Attendance denotes the percentage of classes attended. Regular attendance is hypothesized to have a strong positive influence on CGPA, as it reflects both engagement and discipline.

4. ***Screen Time (hours/day):***

   Screen time captures the daily time spent on screens for non-academic purposes. Excessive screen time may distract from studies and negatively impact CGPA.

Model Overview

The regression model is expressed as:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$

$\text{CGPA} = \beta_0 + \beta_1(\text{Semester}) + \beta_2(\text{Study Hours}) + \beta_3(\text{Attendance}) + \beta_4(\text{Screen Time}) + \epsilon$

Where:

- β_0: Intercept of the regression model.

- X_1, X_2, X_3, X_4: Independent variables, Semester, Study Hours, Attendance, Screen Time, respectively.
- β_1, β_2, β_3, β_4: Regression coefficients for the independent variables.
- ϵ: Error term accounting for unexplained variance.

The model summary yielded an R^2 value of approximately 0.4741, indicating that the predictors collectively explain 47.4% of the variance in CGPA. Detailed results are discussed in the next section.

4. Description, Results, and Discussion

4.1 Description of the Data

The dataset comprises information on 47 students with the following variables: Semester, Study Hours (hours/week), Attendance (%), Screen Time (hours/day), and CGPA. Key observations and summary statistics are:

- Semester:

  Ranges from 3 to 8, with a mean of 3.745. The mode is 3, indicating a concentration of students in earlier semesters.
- Study Hours:

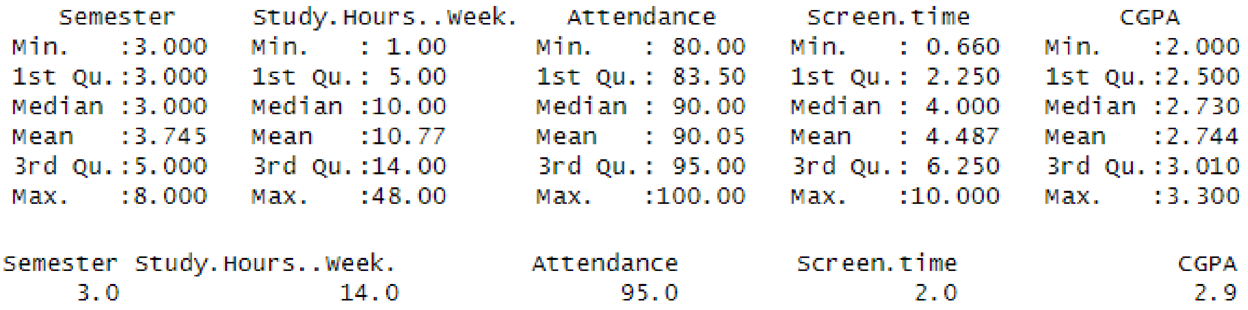
  Averages 10.77 hours/week, with some students reporting as much as 48 hours/week.
- Attendance:

  Averages 90.05%, reflecting high engagement overall.
- Screen Time:

  Averages 4.487 hours/day, with a maximum of 10 hours/day, suggesting variability in screen usage habits.
- CGPA:

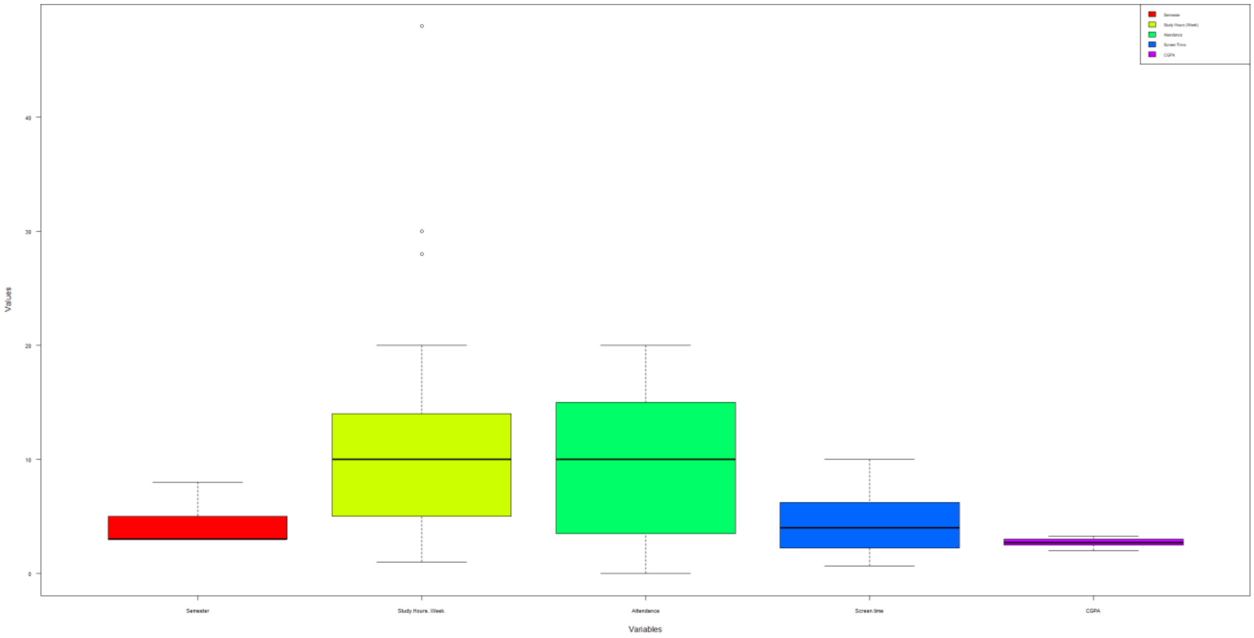
  Ranges from 2.00 to 3.30, with a mean with 2.744.

Summary statistics are as follows:



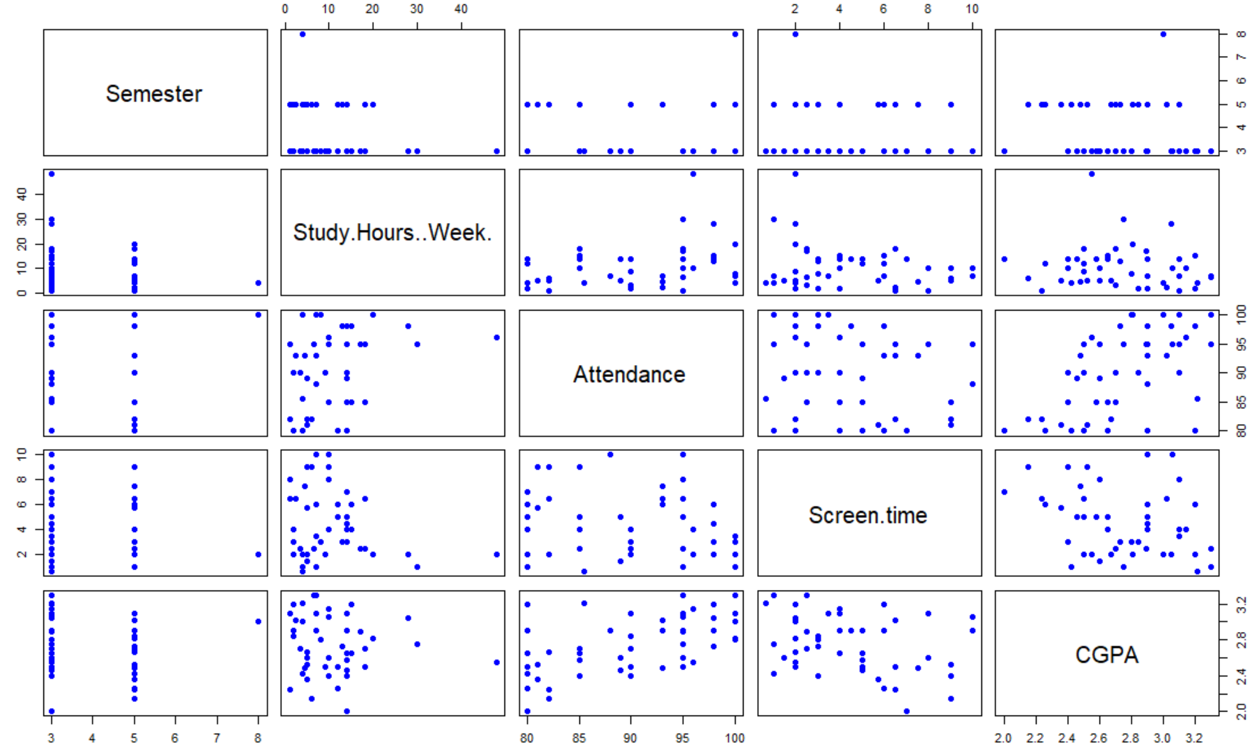
```
    Semester      Study.Hours..Week.   Attendance       Screen.time          CGPA     
 Min.   :3.000   Min.   : 1.00       Min.   : 80.00   Min.   : 0.660   Min.   :2.000  
 1st Qu.:3.000   1st Qu.: 5.00       1st Qu.: 83.50   1st Qu.: 2.250   1st Qu.:2.500  
 Median :3.000   Median :10.00       Median : 90.00   Median : 4.000   Median :2.730  
 Mean   :3.745   Mean   :10.77       Mean   : 90.05   Mean   : 4.487   Mean   :2.744  
 3rd Qu.:5.000   3rd Qu.:14.00       3rd Qu.: 95.00   3rd Qu.: 6.250   3rd Qu.:3.010  
 Max.   :8.000   Max.   :48.00       Max.   :100.00   Max.   :10.000   Max.   :3.300  

Semester Study.Hours..Week.         Attendance        Screen.time               CGPA 
     3.0               14.0               95.0                2.0                2.9 
```

Box plots for all variables:



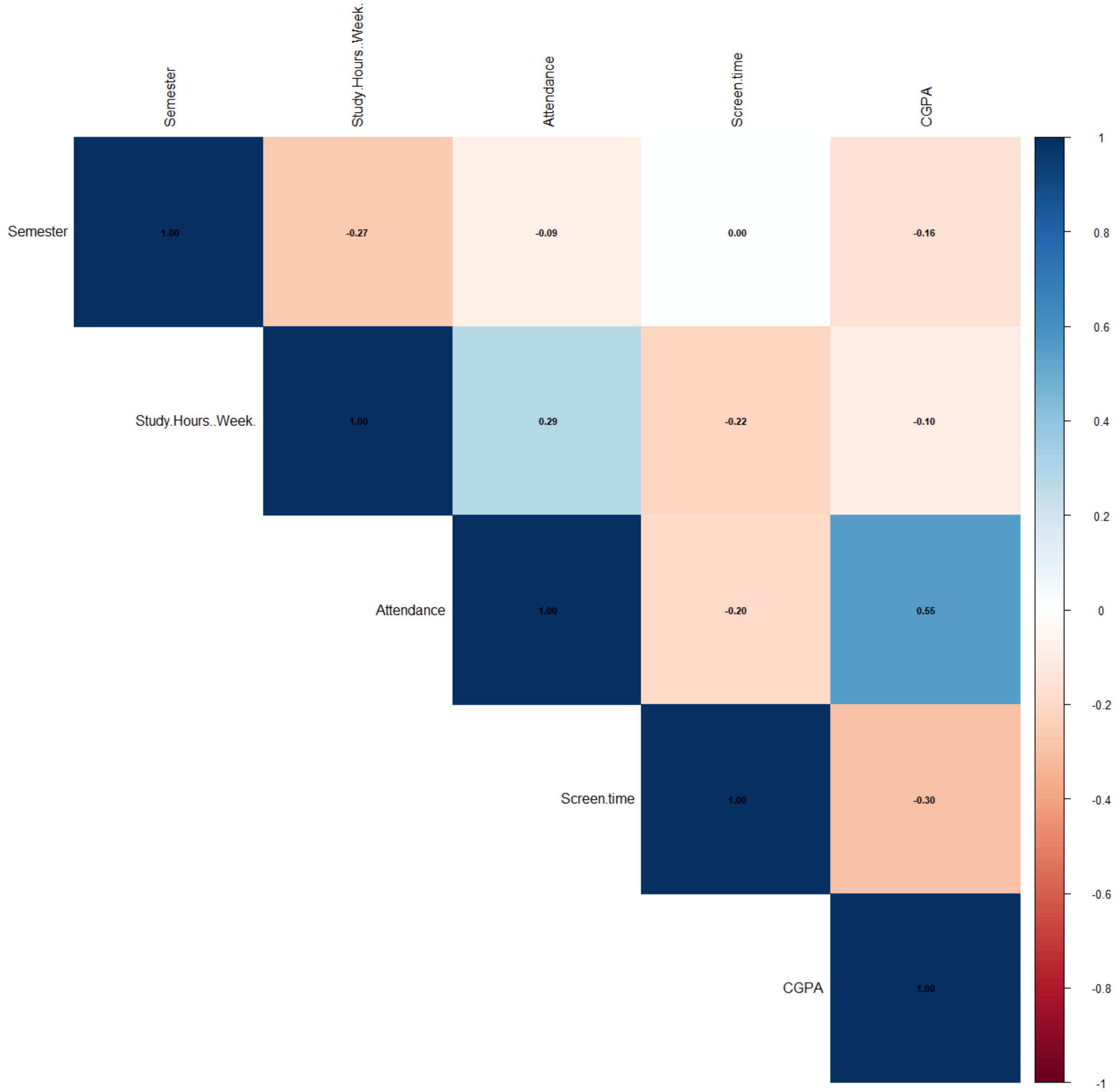
The box plot analysis reveals outliers in Study hours per Week, indicating the presence of exceptionally dedicated students in the dataset. No outliers are observed in the other variables.

Plotting a matrix of scatter plot reveals information about correlation, informally:



From the scatter plot matrix, we focus on the last column, where our target variable, CGPA, is plotted on the x-axis. Examining the intersection between Semester and CGPA, we observe no correlation. Similarly, the intersection between Study Hours and CGPA also shows no discernible correlation. However, the Attendance and CGPA intersection reveals a positive correlation. Lastly, the intersection between Screen Time and CGPA indicates a negative correlation.

A correlation heatmap is included as well (Observe the last column):



A correlation analysis reveals correlation, formally:

- Strong positive correlation between Attendance and CGPA (r = 0.552)
- Weak negative correlation between Screen Time and CGPA (r = -0.297)
- Weak negative correlation between Semester and CGPA (r = -0.159)
- Weak negative correlation between Study Hours and CGPA (r = -0.095)

Low correlation value for Study Hours indicates limited direct influence.

4.2 Results

The regression analysis produced the following prediction equation:

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4$

X_1, X_2, X_3, X_4: Independent variables, Semester, Study Hours, Attendance, Screen Time, respectively.

Where $\hat{\beta}_k$ are predictors, and $\hat{Y}$ is the predicted value of CGPA, shown in the equation below:

CGPA = 0.820 – 0.058 × Semester – 0.014 × Study Hours + 0.027 × Attendance – 0.032 × Screen Time

- **Intercept:** $\hat{\beta}_0 = 0.820$

  The baseline CGPA when all predictors are zero.

- **Semester:** $\hat{\beta}_1 = -0.058$, $p = 0.081$

  Negatively associated with CGPA, though the effect is marginally insignificant. Advanced semesters may introduce greater academic challenges, lowering CGPA.

  CGPA decreases by 0.058 by increasing semester by 1 unit.

- **Study Hours:** $\hat{\beta}_2 = -0.014$, $p = 0.0036$

  Negatively associated with CGPA. Excessive study hours may reflect inefficiencies or overburdening.

  CGPA decreases by 0.014 by increasing study hours by 1 hour per week.

- **Attendance:** $\hat{\beta}_3 = 0.027$, $p < 0.001$

  The most significant positive predictor, indicating that regular attendance strongly enhances academic performance.

  CGPA increases by 0.027 by increasing attendance by 1%.

- **Screen Time:** $\hat{\beta}_4 = -0.032$, $p = 0.028$

  Negatively associated with CGPA, highlighting the detrimental effects of excessive non-academic screen usage.

CGPA decreases by 0.032 by increasing screen time by 1 hour per day.

The Mean Squared Error (MSE), a key measure of model performance, is calculated as the square of the residual standard error (RSE), which is 0.242.

Therefore, the MSE is:

$$\text{MSE} = (0.242)^2 = 0.0585$$

The MSE represents the average squared difference between the observed CGPA values and the predicted CGPA values. A lower MSE indicates a better fit of the model to the data. An MSE of 0.0585 indicates that, on average, the model's predictions deviate from the actual values by about 0.242 points on the CGPA scale.

4.3 Discussion

1. **Key Findings**

   - Attendance emerges as the strongest predictor of CGPA, reinforcing its importance for academic success. Students who attend classes regularly tend to perform better, which emphasizes the need for engagement and discipline.
   - Study hours negatively correlated with CGPA, which may seem counterintuitive. However, this suggests that excessive study time could lead to diminishing returns due to inefficiency, stress, or a lack of proper time management.
   - Screen time also shows a negative correlation with CGPA. This aligns with previous research indicating that excessive non-academic screen usage can be a major distraction, reducing study time and focus.

- Semester is slightly negatively associated with CGPA, likely due to the increasing complexity and workload of courses in later semesters. However, the effect is marginal and not statistically significant, suggesting that other factors may be more influential at this stage.

2. **Model Performance and Implications**

- The MSE value of 0.0585 indicates the average squared difference between the predicted and actual CGPA values. This relatively low value suggests that the model performs reasonably well but also leaves room for improvement. The model explains about 47.4% of the variance in CGPA, meaning that nearly half of the academic performance variance can be accounted for by the selected predictors.
- For students, the model highlights that improving attendance and managing screen time are key strategies for improving academic performance. Students should focus on maximizing their engagement with classes and minimizing distraction from non-academic screen usage.
- For educators, the results suggest that efforts to encourage regular class attendance and reduce distractions in the learning environment can have a significant impact on student success.
- For future research, the model could be expanded by incorporating other factors such as socioeconomic status, extracurricular activities, or personal characteristics like motivation or study strategies, which could potentially improve the model's prediction accuracy.

3. Limitations

   - The study relies on a limited set of predictors, explaining only 47.4% of the variance in CGPA. Additional variables might improve the model's explanatory power.

- The sample size of 47 students may not be large enough to generalize about the broader student population. Larger, more diverse datasets could lead to more robust results.
- The cross-sectional nature of the data limits the ability to draw casual conclusions. Longitudinal studies could provide more insights into how these factors affect CGPA over time.

References

Abu Hasan, N. A., Ahmad, N., & Abdul Razak, N. A. (n.d.). Factors that significantly affect college students' CGPA. Universiti Teknologi MARA.

Akhter, R., & Rahaman, M. S. (2020). Factors affecting academic performance of university students: A study among the students of MBSTU. Journal of Economics and Sustainable Development, 11(20), 16-25.

Erdem, C., Şentürk, İ., & Arslan, C. K. (2007). Factors affecting grade point average of university students. The Empirical Economics Letters, 6(5), 413-422.

Mahmud, S., Jobayer, M. A. A., Salma, N., Mahmud, A., & Tamanna, T. (2023). Online gaming and its effect on academic performance of Bangladeshi university students: A cross-sectional study. Health Science Reports, 6(1).