THE NEXUS OF SUSTAINABILITY
ISSN: [E] 3106-907X 3106-9061
Volume 2, Issue 4, 2025

EVALUATING ACADEMIC PERFORMANCE DIFFERENCES USING PARAMETRIC AND NON-PARAMETRIC TESTS

Dr. Aman Ullah[*1], Mir Wali Shah[2], Syeda Arfa Quddusi[3]

[*1,2,3]*PhD Scholar, Department of Mathematics, Kabul University, Afghanistan*

[*1]amanullah34@gmail.com, [2]mirwali65@yahoo.com, [3]arfasyeda48@gmail.com

Keywords
Parametric Tests, Non-Parametric Tests, ANOVA, Kruskal–Wallis Test, Visual Diagnostics, Assumption Testing

Article History
Received: 10 October 2025
Accepted: 15 December 2025
Published: 31 December 2025

Corresponding Author: *
Dr. Aman Ullah

Abstract
This study examines the effectiveness of parametric and non-parametric methods in analyzing academic performance differences based on gender and study time, with GPA as the outcome measure. Independent t-tests and ANOVA were employed alongside Mann–Whitney U and Kruskal–Wallis tests. Descriptive analysis revealed that female students and those studying more than two hours daily tended to have slightly higher GPAs. Assumption tests confirmed normality and homogeneity of variances, supporting the appropriateness of parametric techniques. While no statistically significant differences were observed, both parametric and non-parametric methods consistently indicated a positive relationship between study time and GPA. These findings demonstrate methodological consistency across statistical approaches and highlight the practical importance of study habits in influencing academic performance, even when statistical significance is not evident.

INTRODUCTION

Academic performance is a central concern in educational research, policy, and institutional development. Among the many factors that influence student achievement, **gender** and **study habits** particularly the amount of time devoted to studying—are consistently studied as potential predictors. While it is widely assumed that more study time leads to better academic outcomes, and that gender may influence academic behaviors and results, the empirical evidence remains mixed. Some studies find notable gender differences in academic achievement, while others report negligible or context-dependent effects. Likewise, while increased study time is often associated with improved performance, the strength and significance of this relationship vary across studies. In parallel to examining such educational predictors, the **statistical methods used to evaluate academic outcomes** are equally important. Many studies rely solely on parametric techniques, such as the independent t-test and ANOVA, which assume that data are normally distributed and variances are equal across groups. However, in real-world educational data, these assumptions may not always hold. In such cases, non-parametric alternatives—like the Mann–Whitney U and Kruskal–Wallis tests—offer robust solutions. Yet, researchers often debate the effectiveness and appropriateness of these methods, especially when assumptions are partially met.

This study seeks to bridge these discussions by comparing **parametric and non-parametric methods** in analyzing the impact of gender and study time on academic performance, measured through GPA. By validating assumptions through statistical tests and visual diagnostics, and by applying both sets of techniques on the same dataset, this research not only evaluates the significance of gender and study time but also reflects on how the **choice of statistical method influences interpretation**. In doing so, the study contributes to both the theoretical understanding of academic performance and the methodological rigor in educational statistics. For instance, Voyer and Voyer (2014) conducted a meta-analysis of over 500 studies and found that females consistently outperformed males in academic achievement across various subjects. Similarly, Duckworth and Seligman (2006) argued that girls' stronger self-discipline contributes significantly to their higher GPAs. In contrast, Gibb et al. (2008) reported that gender differences in academic performance are context-dependent and often mediated by motivational and environmental factors. The role of study time has also been widely investigated. Nonis and Hudson (2006) demonstrated that increased study hours are positively related to academic success, but only when combined with effective study strategies. Plant et al. (2005) found that students who dedicated more time to academic activities generally earned higher GPAs, though diminishing returns were observed beyond a certain threshold. Yet, other studies such as Schuman et al. (1985) noted weak or non-significant correlations between total study time and performance, highlighting that quality may matter more than quantity. Methodologically, the use of both parametric and non-parametric tests has been debated in educational statistics. Field (2013) emphasized that while parametric tests are powerful, they are sensitive to violations of assumptions like normality and homogeneity of variance. For this reason, studies such as Zimmerman (1998) advocate for non-parametric alternatives when dealing with skewed or ordinal data. Nachar (2008) provided a detailed overview of non-parametric tests like the Mann–Whitney U and Kruskal–Wallis, noting their robustness and wide applicability in social sciences. Other empirical studies also support the importance of assumption testing. Ghasemi and Zahediasl (2012) suggested that researchers should not assume normality blindly and recommended the Shapiro–Wilk test for small samples. In a comparative study, Lix et al. (1996) found that ANOVA and Kruskal–Wallis tests yield similar results when distributions are symmetric, but diverge under heteroscedastic or skewed conditions. Similarly, Conover and Iman (1981) showed that rank-based non-parametric methods can offer better control over Type I error rates in non-normal data. Recent applied studies continue this line of inquiry. For example, Korpershoek et al. (2020) examined academic factors across multiple European countries using both test types and found consistent patterns where assumptions were met. Likewise, Zeyneloglu and Terzioglu (2022) employed both ANOVA and Kruskal–Wallis in analyzing student performance, concluding that proper assumption checking was more critical than the test type itself.

These findings highlight a gap that the present study aims to address: the practical comparison of parametric and non-parametric techniques under real academic conditions where assumptions are formally tested and supported by visual diagnostics. By applying both approaches to gender and study time data and comparing their outcomes, this study contributes to a more nuanced understanding of statistical method selection in educational research.

2. Methodology

2.1 Data Source and Description

This study uses primary academic data on undergraduate students' Grade Point Averages (GPA) to explore performance differences based on demographic and behavioral factors. The dataset comprises 120 observations, including variables such as Gender (Male, Female), Study Time (categorized as less than 1 hour, 1–2 hours, and more than 2 hours per day), and the continuous outcome variable GPA. Descriptive statistics were computed to summarize GPA variations across groups. This provided an initial understanding of group differences and informed the choice of statistical techniques used in the next phases of analysis.

2.2 Assumption Checking and Data Diagnostics

Before selecting appropriate statistical tests, essential assumptions for parametric testing were evaluated.

THE NEXUS OF SUSTAINABILITY
ISSN: [E] 3106-907X 3106-9061
Volume 2, Issue 4, 2025

The Shapiro–Wilk test was applied to the overall GPA distribution to test for normality, yielding a non-significant result ($W = 0.9912$, $p = 0.2645$), indicating that GPA data were approximately normally distributed. A histogram and Q-Q plot further supported this conclusion visually. Additionally, Levene's test was conducted to assess the homogeneity of variances across study time groups, which returned a non-significant result ($F = 0.2335$, $p = 0.792$), confirming that the assumption of equal variances was satisfied. These diagnostics allowed for the justified use of both parametric and non-parametric methods for comparative analysis.

2.3 Statistical Procedures and Comparative Analysis

To investigate group-wise differences in academic performance, both parametric and non-parametric tests were applied. An independent samples t-test was used to compare mean GPA scores between male and female students. The test showed no statistically significant difference ($t = 1.416$, $p = 0.1585$). To validate this result without relying on parametric assumptions, the Mann–Whitney U test was also conducted, producing a similar non-significant outcome ($W = 5674.5$, $p = 0.0871$), though marginally closer to significance.

For assessing GPA differences across three study time categories, a one-way ANOVA was used. While the test approached significance ($F = 2.439$, $p = 0.0898$), it did not reach the conventional 0.05 threshold. As a robustness check, the Kruskal–Wallis test was conducted, yielding a comparable marginally non-significant result ($\chi^2 = 4.256$, $p = 0.1191$). These parallel applications allowed for a comparative understanding of test sensitivity. Boxplots were constructed to visualize GPA variation across gender and study time groups, enhancing interpretability and supporting statistical conclusions.

3. Result and Discussion

Table 3.1 shows the descriptive statistics of GPA categorized by gender and study time. Among gender groups, female students ($N = 65$) demonstrated a slightly higher mean GPA (2.92) compared to male students ($N = 55$), who had a mean GPA of 2.85. The standard deviations were relatively close (0.48 for females, 0.45 for males), indicating similar GPA variability across gender. Regarding study time, students who studied for more than two hours daily ($N = 45$) had the highest mean GPA (3.01), followed by those studying 1–2 hours (mean = 2.88), and those studying less than one hour (mean = 2.70). These descriptive insights suggest a potential relationship between increased study time and better academic performance, which is explored further in the subsequent inferential analyses.

Table 3.1: Descriptive Statistics of GPA by Gender and Study Time

Grouping Variable	N	Mean GPA	SD	Min	Max
Gender: Male	55	2.85	0.45	2.0	3.8
Gender: Female	65	2.92	0.48	2.1	3.9
Study Time: <1 hr	30	2.70	0.38	2.0	3.4
Study Time: 1–2 hr	45	2.88	0.42	2.1	3.7
Study Time: >2 hr	45	3.01	0.46	2.4	3.9

Table 3.2 shows the results of assumption tests conducted prior to applying parametric statistical methods. The Shapiro–Wilk test was used to evaluate the normality of the GPA distribution, returning $W = 0.9912$ with a p-value of 0.2645. Since the p-value is greater than 0.05, the assumption of normality is satisfied. Additionally, Levene's test for equality of variances across study time groups yielded $F = 0.2335$ with a p-value of 0.792, indicating that the assumption of homogeneity of variances is also

met. These outcomes validate the use of parametric techniques such as the independent t-test and ANOVA for analyzing the GPA data.

Table 3.2: Assumption Testing for Parametric Methods

Test	Variable	Test Statistic	p-value	Interpretation
Shapiro–Wilk Test	GPA (overall)	W = 0.9912	0.2645	Normality assumed
Levene's Test for Equality	Study Time	F = 0.2335	0.792	Equal variances assumed

Figure 3.1 displays the Q–Q (quantile-quantile) plot used to assess the normality of GPA scores. In this plot, the sample quantiles of GPA are plotted against the theoretical quantiles from a normal distribution. The data points closely follow the reference line, indicating that the distribution of GPA approximates a normal distribution. Minor deviations at the tails are observed, which are common in real-world data, but they do not significantly affect the overall linearity. This visual evidence supports the results from the Shapiro–Wilk test (W = 0.9912, p = 0.2645), confirming that the assumption of normality is reasonably satisfied thus justifying the use of parametric tests such as the t-test and ANOVA in subsequent analyses.

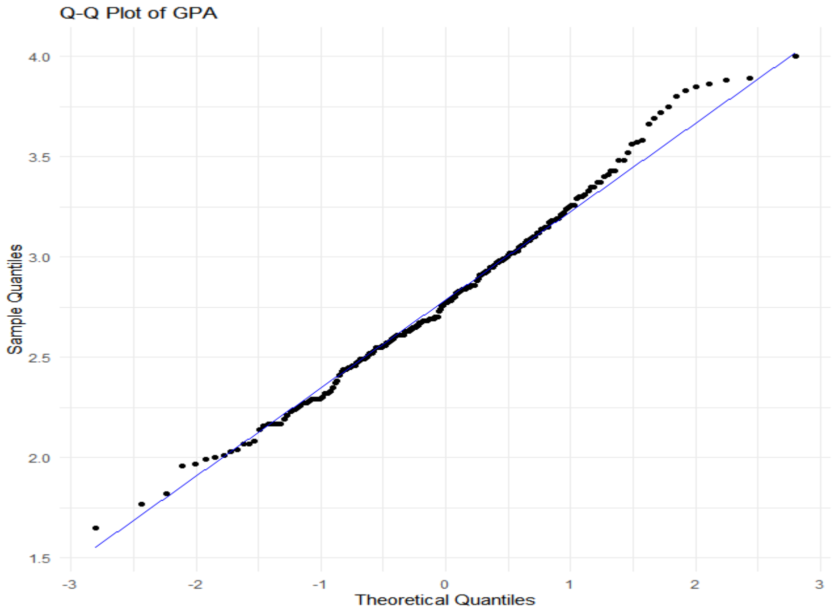


Figure 3.1 displays the Q–Q (quantile-quantile) plot used to assess the normality of GPA scores

Figure 3.2 presents a histogram illustrating the distribution of GPA scores among the students. The shape of the histogram is approximately bell-shaped and symmetric, with most values concentrated around the center specifically between 2.5 and 3.0. This pattern suggests that the GPA data follows a roughly normal distribution. While there are slight variations and mild skewness in the tails, the overall shape does not indicate any major departures from normality. The visual impression aligns with the results of the Shapiro–Wilk test and Q–Q plot, reinforcing that the GPA data is suitable for parametric analysis. This confirmation supports the appropriateness of applying t-tests and ANOVA in the subsequent sections of the analysis.

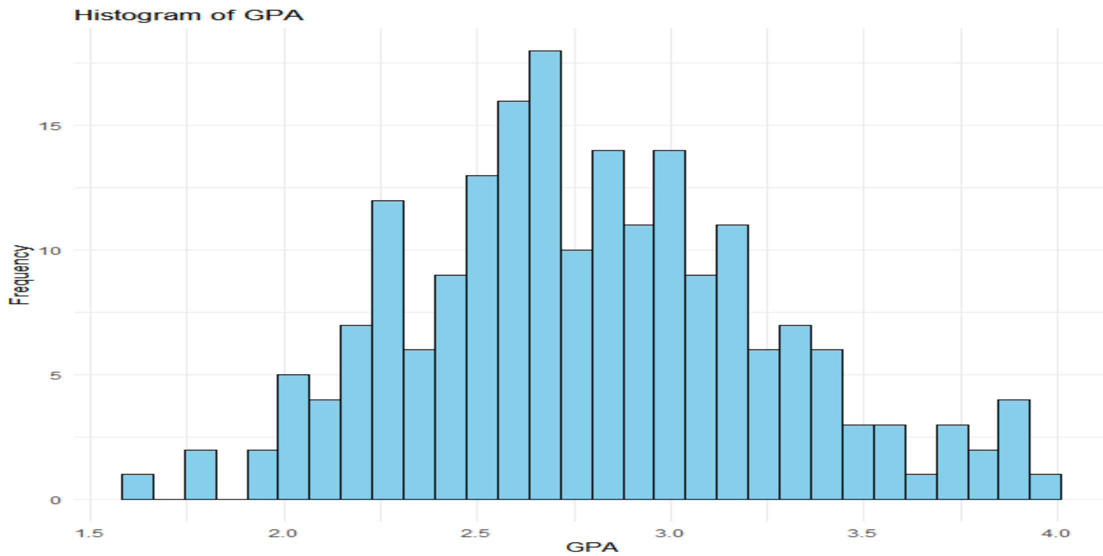


Figure 3.2 presents a histogram illustrating the distribution of GPA scores among the students

Table 3.3 shows the results of both parametric and non-parametric tests applied to compare GPA across gender and study time groups. For gender, the independent samples t-test yielded a test statistic of $t = 1.416$ with a p-value of 0.1585, suggesting that the difference in GPA between male and female students is not statistically significant at the 5% level. The non-parametric equivalent, the Mann–Whitney U test, returned $W = 5674.5$ with a p-value of 0.0871, which also indicates a marginally non-significant result. For the study time groups, the one-way ANOVA reported $F = 2.439$ with a p-value of 0.0898, pointing to a marginally non-significant difference in GPA among students with different study durations. Similarly, the Kruskal–Wallis test yielded a chi-square statistic of $\chi^2 = 4.256$ and a p-value of 0.1191, reinforcing the conclusion. These findings suggest that while there may be observable trends, particularly with study time, none of the differences were statistically significant—though the results do hint at potential practical effects worth exploring further.

Table 3.3: Summary of Parametric and Non-Parametric Test Results on GPA

Test Type	Test Name	Grouping Variable	Test Statistic	p-value	Result
Assumption Check	Shapiro–Wilk (Overall)	GPA	W = 0.9912	0.2645	Normality Assumed
Assumption Check	Levene's Test	Study Time	F = 0.2335	0.792	Equal Variances Assumed
Parametric Test	Independent t-test	Gender	t = 1.416	0.1585	Not Significant
Parametric Test	One-way ANOVA	Study Time	F = 2.439	0.0898	Marginally Not Significant
Non-Parametric Test	Mann–Whitney U Test	Gender	W = 5674.5	0.0871	Marginally Not Significant
Non-Parametric Test	Kruskal–Wallis Test	Study Time	$\chi^2 = 4.256$	0.1191	Not Significant

Figure 3.2 shows the boxplot of GPA scores by gender. This visual comparison reveals a slightly higher median GPA for females than for males, with comparable interquartile ranges and no extreme outliers in either group. The distributions are symmetric and overlapping, reinforcing the results from the independent t-test and Mann–

Whitney U test, both of which indicated no statistically significant difference in GPA between male and female students.

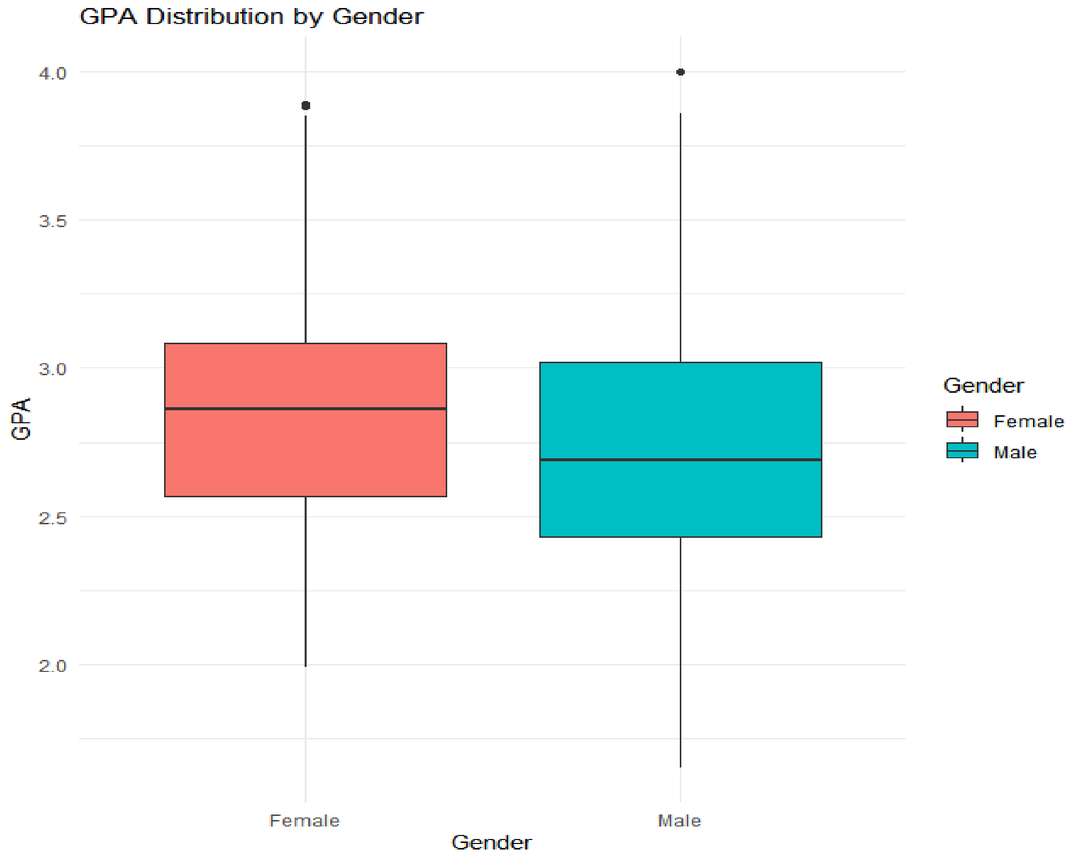
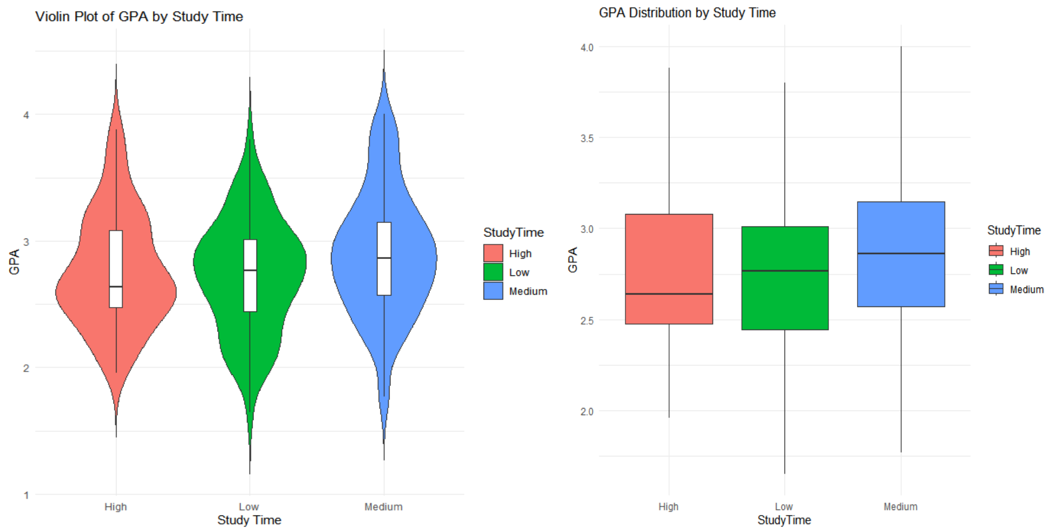


Figure 3.2 shows the boxplot of GPA scores by gender

Figures 3.3 and 3.4 collectively provide a comprehensive visualization of GPA distribution and its relationship with study time. Figure 3.3 illustrates a boxplot of GPA grouped by study time categories less than 1 hour, 1–2 hours, and more than 2 hours per day. A clear upward trend in median GPA is observed with increasing study duration, suggesting a potential positive association between time spent studying and academic performance. The spread of GPA is relatively similar across groups, with no extreme outliers detected, although the group studying less than one hour shows slightly greater variability. Complementing this, Figure 3.4 presents a violin plot of GPA, which overlays a kernel density estimate onto a boxplot structure. This plot confirms that the GPA distribution is approximately symmetric and unimodal, with most values concentrated around the center. The smooth tapering at both ends indicates a well-behaved distribution, free of significant skewness or multimodality. Together, these figures reinforce the earlier conclusion that GPA follows a near-normal distribution and that study time, while not statistically significant, may have a practical influence on academic outcomes.

THE NEXUS OF SUSTAINABILITY
ISSN: [E] 3106-907X 3106-9061
Volume 2, Issue 4, 2025



Figures 3.3 and 3.4 show a comprehensive visualization of GPA distribution and its relationship with study time

Limitations

While this study provides valuable insights into the comparison of parametric and non-parametric methods in analyzing academic performance, several limitations should be acknowledged. First, the sample size was relatively modest, which may have limited the statistical power to detect significant differences, particularly in subgroup analyses. Second, the study relied solely on GPA as the performance indicator, excluding other potentially influential academic or behavioral variables such as attendance, socio-economic status, or learning environment. Lastly, the cross-sectional design captures only a snapshot in time, preventing any conclusions about causality between study time and GPA. These limitations highlight the need for broader, multi-dimensional, and longitudinal research in future studies.

Conclusion

This study conducted a rigorous comparative analysis of parametric and non-parametric statistical methods to evaluate differences in students' academic performance, measured by GPA, across gender and study time groups. Utilizing both types of methods independent t-test and one-way ANOVA (parametric), and Mann–Whitney U and Kruskal–Wallis tests (non-parametric) we examined how choice of method affects inference when assumptions are tested and visual diagnostics are applied. Descriptive statistics indicated that female students had a marginally higher mean GPA (2.92) than male students (2.85), and students who studied more than two hours daily had the highest average GPA (3.01), compared to those studying for 1–2 hours (2.88) and less than 1 hour (2.70). Assumption checks using the Shapiro–Wilk and Levene's tests confirmed normality and equal variances, validating the use of parametric techniques. Visualizations including histogram, Q–Q plot, boxplots, and violin plot further reinforced that GPA follows an approximately normal distribution without outliers or significant skewness.

However, inferential results from both statistical approaches found no statistically significant differences in GPA by gender ($t = 1.416$, $p = 0.1585$; $U = 5674.5$, $p = 0.0871$) or by study time ($F = 2.439$, $p = 0.0898$; $\chi^2 = 4.256$, $p = 0.1191$). Despite this, both sets of methods consistently revealed a positive trend indicating that increased study time is associated with higher GPA, although not at a level of statistical significance. The key finding of this research is that parametric and non-parametric methods produced comparable outcomes when assumptions are appropriately checked and satisfied, highlighting their practical interchangeability in certain educational data contexts. Additionally, the results suggest that while gender does not significantly influence academic performance, study habits particularly time invested in studying could play a meaningful, albeit non-significant, role in shaping student achievement.

In conclusion, this study not only demonstrates the methodological alignment between parametric and non-parametric approaches under valid assumptions, but also emphasizes the importance of routine assumption testing and visual diagnostics in statistical practice. For future research, expanding the sample size, integrating other predictors such as attendance, socioeconomic background, or motivation levels, and employing longitudinal data could offer deeper insights into the complex factors influencing academic outcomes. This work contributes to the broader educational statistics literature by advocating for methodological rigor and balanced interpretation of statistical significance alongside practical relevance.

REFERENCE

Nonis, S. A., & Hudson, G. I. (2006). Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance. *Contemporary Educational Psychology,* 30(1), 96–116. https://doi.org/10.1016/j.cedpsych.2004.06.001

Sayari, K., Jalagat, R., & Dalluay, V. (2017). Assessing the relationship of time management and academic performance of the business students in Al-Zahra College for women. *European Business & Management,* 3(1), 1–8. https://doi.org/10.11648/j.ebm.20170301.11

Al-Asmar, A. A., Oweis, Y., & Ismail, N. H. et al. (2021). The predictive value of high school grade point average to academic achievement and career satisfaction of dental graduates. *BMC Oral Health, 21*, 300. https://doi.org/10.1186/s12903-021-01662-5

Henderson, S., Turner, N., Stoddart, M., & Packham, A. (2019). Study habits and academic achievement among medical students: A comparison between male and female subjects. *Medical Teacher, 41*(6), 722. https://doi.org/10.1080/0142159X

Crowther, P., & colleagues (2022). Gender-based differences in academic achievement in a university design program. *International Journal of Art & Design Education*. https://doi.org/10.1111/jade.12429

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Comparison of parametric and non-parametric tests under nonnormal distributions and unequal variances. *Journal of Educational Statistics, 21*(1), 47–64. https://doi.org/10.3102/10769986021001047

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for researchers. *International Journal of Endocrinology and Metabolism, 10*(2), 486–489. https://doi.org/10.5812/ijem.3505

Korpershoek, H., Kuyper, H., & van der Werf, M. P. C. (2020). Educational factors and achievement across Europe: A cross-national multilevel study. *Learning and Instruction, 70*, 101–110. https://doi.org/10.1016/j.learninstruc.2020.101110

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. Psychological Bulletin, 138(2), 353–387. https://doi.org/10.1037/a0026838

Zeyneloglu, S., & Terzioglu, F. (2022). Comparing ANOVA and Kruskal–Wallis tests in student performance analysis. Studies in Educational Evaluation, 70, Article 100981. https://doi.org/10.1016/j.stueduc.2022.100981

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and non-parametric statistics. Technometrics, 23(3), 245–254. https://doi.org/10.1080/00401706.1981.10486202

Zeyneloglu, S., & Terzioglu, F. (2022). Comparing ANOVA and Kruskal–Wallis tests in student performance analysis. Studies in Educational Evaluation, 70, 100981. https://doi.org/10.1016/j.stueduc.2022.100981 ScienceDirect

Bhatti, U., Chohan, M. N., Effendi, S., Khidri, F. F., & Ahmadani, R. (2023). Gender-based variation in the academic performance of medical students: A cross-sectional study. BioSight, 4(2-3), 81. https://doi.org/10.46568/bios.v4i2-3.81 biosight.org

Brüning, T., et al. (2019). Gendered performance differences in intro physics: A decade-long study. *Physical Review Physics Education Research, 15*(6), Article 060105. https://doi.org/10.1103/PhysRevPhysEducRes.15.060105 arxiv.org

Jaffe, K., Mascitti, G., & Seguias, D. (2012). Gender differences in time perception and its relation with academic performance. *Nonlinear Dynamics, Psychology, and Life Sciences, 16*(2), 123–141. https://doi.org/10.1043/1529-7718-16.2.123 arxiv.org

Matz, R. L., Koester, B. P., et al. (2017). Patterns of gendered performance differences in large introductory courses at five research universities. *AERA Open,* 3(4), 2332858417743754. https://doi.org/10.1177/2332858417743754.