

Determinants of Student Test Scores: Academic, Family, and Learning Environment Factors a Multiple Linear Regression Analysis

Rena Nainggolan¹, Fenina A.T. Tobing²

¹ Komputerisasi Akuntansi Study Program, Universitas Methodist Indoensia

² Informatics Study Program, Universitas Multimedia Nusantara, Tangerang, Indonesia

¹ renanain66olan@gmail.com

Accepted 28 May 2026

Approved 17 July 2026

Abstract— This study aims to analyze the influence of academic, family, and learning environment factors on student exam scores. The study employs a quantitative approach using secondary data in the form of the Student Performance Factors dataset comprising 6,607 observations. The dependent variable in this study is student exam scores (Exam Score), while the independent variables include study hours (Hours Studied), attendance (Attendance), previous scores (Previous Scores), tutoring sessions (Tutoring Sessions), parental involvement (Parental Involvement), and access to learning resources (Access to Resources). The analytical techniques used were descriptive statistics and multiple linear regression. The results indicate that the regression model is statistically significant simultaneously with $F=1,601$ and $p < 0.001$, and explains 66.0% of the variance in students' exam scores ($R^2 = 0.660$). Partially, study hours, attendance, previous scores, and tutoring sessions have a positive and significant effect on test scores. Additionally, students with higher levels of parental involvement and access to learning resources tend to achieve better test scores. These findings indicate that students' academic achievement is not solely determined by individual study effort but is also influenced by family support and the availability of an adequate learning environment.

Index Terms—Determinants of Exam Scores; Learning Environment; Multiple Linear Regression; Parental Involvement; Student Academic Performance.

I. INTRODUCTION

Exam scores are a key indicator in assessing students' academic achievement [1]. Good learning outcomes not only reflect mastery of the material but also demonstrate the effectiveness of the learning process occurring both at school and at home. In practice, students' academic achievement is influenced by various interacting factors, such as study habits, attendance rates, family support, and learning environment conditions[1].

Academic factors such as study intensity, discipline in attending school, prior learning experiences, and participation in tutoring are often considered direct determinants of student success [1]. The higher a student's engagement in the learning process[2], the

greater their chances of achieving better grades. However, academic factors are not the sole explanatory. The family environment [3], particularly parental involvement [4], also plays a crucial role in shaping students, motivation, study habits, and academic supervision.

Additionally, a supportive learning environment such as the availability of adequate learning resources can help students understand the material more effectively [5]. Students with better access to instructional materials, supporting tools, or learning facilities tend to develop academic understanding more easily compared to those with limited access [6]. Therefore, analyses of academic achievement must consider students holistically, not only from an individual perspective but also in terms of the external support they receive [7]. Discussing the use of learning analytics and educational data mining to predict student achievement. It's an educational analytics approach, not linear regression [8]. Given these conditions, this study was conducted to examine the determinants of students' exam scores through academic, family, and learning environment factors using multiple linear regression analysis [1][3]. This study is expected to provide an empirical overview of the factors most influential on exam scores, thereby serving as a basis for improving the quality of learning and educational support for students [9]. Previously, the study focused on a single aspect of academic performance, such as learning, individual, or learning environment. Other studies use data-based education or multilevel education to predict academic performance [10]. However, a joint study that examines academic, familial, and learning environment factors using a berganda regression on a large student skala dataset consisting of more than 6,000 observations is still unreliable. Because of this, this study integrates these three dimensions into a single regression model in order to identify the factors that have the greatest impact on student test scores [8].

II. RESEARCH METHODS

This study uses a quantitative eksplanatory approach to analyze the impact of academic, familial, and learning environment factors on students' exam scores (Exam_Score). The first phase of research is completed by identifying research challenges and determining research objectives based on observed phenomena [11].

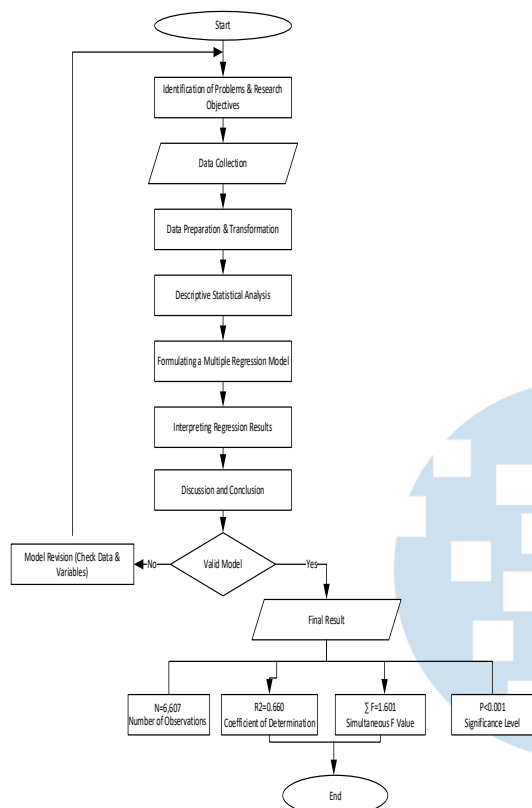


Fig 1. Research Methodology Flow

The next step involves gathering data using a publicly available dataset under Student Performance Factors. This dataset consists of 6.607 observations with a dependent variable called Exam_Score, while independent variables include academic, familial, and learning environment factors.

Following data collection, data transformation and persiapan are carried out. This article discusses data consistency, outliers, and missing values. The categorical variable is then transformed into a dummy variable so that it may be used in regression analysis. "High" is used as a reference category for the variables Parental Involvement and Access to Resources during the coding process.

Next, a deskriptif statistical analysis is performed to illustrate the characteristics of the research data. The analysis results show that the rata-rata Exam Score is 67.24 with a standard deviation of around 3.89. In addition, rata-rata Hours Studied was approximately 19.98 hours, while rata-rata Attendance reached 79.98%. This analysis also compares the rata-rata ujian based on each category of the research variables.

The next step is to use the linear regression model to increase the influence of independent variables on student test. The regression model that is used can be described as follows:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + e$$

In the aforementioned model, variables X_1 – X_4 are continuous academic variables, while variables X_5 – X_6 are dummy variables that represent family and learning environment factors.

Subsequently, the hypothesis is tested using uji simultan (Uji F) and uji parsial (Uji t). The uji F result is $F = 1.601$ with $p < 0.001$, indicating a fully significant regression model. Conversely, the results of the uji t indicate that every independent variable has a p-value of less than 0.001, meaning that every variable has a significant impact on the student's. The next step is to understand the results of the regression. The coefficient of determination (R^2), which is approximately 0.660, indicates that the model can explain 66.0% of the variation in student ujian, although other factors outside the study model have an impact. The variable Tutoring Sessions has a coefficient of 0.504, making it a factor with the most positive impact on the increase in ujian. Conversely, the category Low Access_to_Resources has a coefficient of 2,037, which indicates a greater negative impact on student achievement.

The study's findings indicate that academic factors are the most powerful and reliable predictors of students' learning performance[12]. In addition, academic performance is affected in a multifaceted way by the combination of academic, familial, and learning environment factors .

Based on the aforementioned results, recommendations that might be made include increasing learning discipline, improving student kehadiran, increasing access to learning resources, and increasing individual participation in the educational process [13].

This study employs a quantitative approach with an explanatory research design, as it aims to explain the influence of several independent variables on a single dependent variable [7]. The data used are secondary data from the Student Performance Factors dataset, comprising 6,607 students.

The factors considered in this investigation are shown in Table 1. Exam score is the dependent variable, and the independent factors are two categorical variables (Parental Involvement and Access to Resources) and four numerical variables (Hours Studied, Attendance, Previous Scores, and Tutoring Sessions). Prior to regression analysis, the categorical data were converted into dummy variables.

The results of this investigation are mostly in line with earlier studies. Consistent with the current findings, [1] found that attendance and study habits have a major impact on students' academic progress.

Similarly, using multiple linear regression, [7] discovered that academic-related factors are among the best indicators of exam achievement. Additionally, the results of [3] who highlighted the significance of family support in enhancing kids' learning outcomes, are consistent with the substantial impact of parental participation seen in our study. In contrast to other research, this study uses a reasonably large dataset of 6,607 observations to integrate academic, family, and learning environment elements into a single regression model, offering a more thorough understanding of the factors influencing students' examination performance.

TABLE I. RESEARCH VARIABLES AND MEASUREMENT

Variable	Symbol	Type	Scale	Description
Exam Score	Y	Dependent	Numerical	Student final examination score
Hours Studied	X1	Independent	Numerical	Number of study hours per week
Attendance	X2	Independent	Percentage	Student attendance rate (%)
Previous Scores	X3	Independent	Numerical	Previous academic achievement score
Tutoring Sessions	X4	Independent	Numerical	Number of tutoring sessions attended
Parental Involvement	X5	Independent	Categorical	High, Moderate, Low
Access to Resources	X6	Independent	Categorical	High, Moderate, Low

A. Data Preprocessing

Preprocessing makes sure the dataset meets the multiple linear regression assumptions. The "High" category was used as the reference level when applying dummy variable encoding to categorical variables. Data Preprocessing Procedure the accuracy and dependability of the model are directly impacted by the quality of the input data, data preparation is an essential stage in regression analysis. The dataset was put through a number of preparation steps to make sure it satisfied the statistical presumptions needed for analysis before the multiple linear regression model was developed. These processes comprised categorical variable encoding, data transformation, outlier detection, missing value identification, duplicate checking, and data quality inspection. Specifically, the High category was used as the reference level to convert the variables Parental Involvement and Access to Resources into dummy variables.

Based on earlier research and the goals of this study, a conceptual framework was created to give a clear picture of the relationships between the research

variables. The framework outlines the independent variables academic, familial, and learning environment that are thought to have an impact on children's test performance. The multiple linear regression model employed in this study was built using this conceptual framework. Figure 2 displays the suggested research framework.

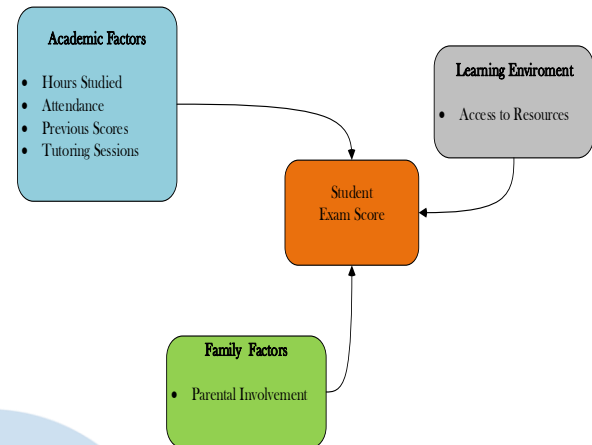


Fig 2. Conceptual Framework of the Proposed Research

A conceptual framework was developed to provide a clear picture of the interactions between the research variables based on previous research and the objectives of this study. The framework describes the independent factors academic, familial, and learning environment that are believed to affect kids' test scores. This conceptual framework was used to construct the multiple linear regression model used in this investigation. The proposed research framework is shown in Figure 2.

TABLE II. DATA PREPROCESSING PROCEDURE

Step	Activity	Purpose
Data Collection	Student Performance Factors Dataset	Obtain research data
Data Cleaning	Duplicate checking	Remove duplicate observations
Missing Value Inspection	Check missing values	Ensure data completeness
Outlier Detection	Boxplot and IQR inspection	Reduce abnormal observations
Data Transformation	Dummy variable encoding	Convert categorical variables
Feature Selection	Select six predictor variables	Improve model relevance
Statistical Assumption Test	Normality, multicollinearity, homoscedasticity	Ensure regression assumptions
Model Development	Multiple Linear Regression	Predict student exam score
Model Evaluation	R ² , F-test, t-test	Evaluate model performance

The analytical approach used in this study is also demonstrated by the conceptual framework. The

multiple linear regression model was used to estimate the contribution of each independent variable to the exam scores of the students. The t-test was used to determine the statistical significance of each predictor, and the F-test and the coefficient of determination (R^2) were used to evaluate the overall model performance. Through statistical analysis, this analytical process guarantees the empirical validation of the links suggested in the conceptual framework.

B. Regression Model Parameters

A number of model evaluation parameters were examined in order to assess the suggested multiple linear regression model's overall performance. These parameters give information about the regression model's predictive power, statistical significance, and goodness-of-fit. Table 3 provides a summary of the evaluation's findings.

TABLE III. REGRESSION MODEL PARAMETERS

Parameter	Value
Number of observations	6,607
Number of predictors	6
Regression model	Multiple Linear Regression
Dependent variable	Exam Score
R^2	0.660
Adjusted R^2	(isi sesuai output software)
F-statistic	1601
Significance (F-test)	$p < 0.001$
Significant predictors	6 variables

Six independent variables and 6,607 observations were used to create the multiple linear regression model, as shown in Table 3. With a coefficient of determination (R^2) of 0.660, the model's predictor variables account for about 66.0% of the variation in students' exam scores. Additionally, the regression model is statistically significant and appropriate for understanding the association between academic, family, and learning environment characteristics and student academic performance, as confirmed by the F-statistic of 1601 with a significance level of $p < 0.001$. These findings show that the suggested regression model fits the data well generally and can be trusted to evaluate how the chosen factors affect students' test scores.

The dependent variable in this study is student exam scores (Exam Score). The independent variables used consist of six main variables, namely study hours (Hours Studied), attendance (Attendance), previous scores (Previous Scores), tutoring sessions (Tutoring Sessions), parental involvement (Parental Involvement), and access to learning resources (Access to Resources). The variables for parental involvement and access to learning resources are categorical [8]; therefore, in the analysis, they were converted into

dummy variables with the "high" category as the reference category.

Data analysis was conducted in two stages. The first stage was a descriptive analysis to describe the characteristics of the data[9]. The second stage was a multiple linear regression analysis to test the simultaneous and partial effects of the independent variables on students' test scores. The regression model used can be formulated as follows [14] [15]:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + e$$

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + e$$

where:

1. Y = student's test score
2. X_1 = study hours
3. X_2 = attendance
4. X_3 = previous score
5. X_4 = tutoring sessions
6. X_5 = parental involvement
7. X_6 = access to learning resources
8. a = constant
9. e = error

Hypothesis testing was conducted using a t-test to examine the partial effects of each variable and an F test to examine the simultaneous effects of all independent variables on students' test scores [16].

III. RESEARCH RESULTS

The results of the descriptive analysis show that the average student exam score was 67.24 with a standard deviation of 3.89. The average number of study hours was 19.98 hours, the average attendance rate was 79.98%, the average previous score was 75.07, and the average number of tutoring sessions was 1.49 sessions.

Categorically, students with high parental involvement had an average test score of 68.09, higher than the moderate category at 67.10 and the low category at 66.36. A similar pattern was also observed regarding access to learning resources, where students with high access had an average score of 68.09, the moderate category 67.13, and the low category 66.20. These findings provide initial indications that family support and the learning environment are associated with academic achievement.

The results of multiple linear regression indicate that the research model is simultaneously significant with an F-value of 1601 and $p < 0.001$. The coefficient of determination, $R^2 = 0.660$, indicates that 66.0% of the variation in students' exam scores can be explained by the six independent variables in the model, while the remaining 34.0% is explained by other factors outside the model.

Partially, all the main variables in the model have a significant effect on students' exam scores. A summary of the regression results is as follows:

TABLE IV. SUMMARY OF THE REGRESSION

Variable	Coefficient	Significance
Study hours	0.294	< 0.001
Attendance	0.199	< 0.001
Previous score	0.048	< 0.001
Tutoring session	0.504	< 0.001
Low parental involvement	-1.996	< 0.001
Moderate parental involvement	-1.035	< 0.001
Low access to learning resources	-2.037	< 0.001
Moderate access to learning resources	-0.943	< 0.001

Based on the table, study hours have a positive effect on test scores. This means that every additional hour of study tends to increase test scores by 0.294 points, assuming other variables remain constant. Attendance also has a positive effect, where each one-point increase in attendance raises the exam score by approximately 0.199 points. This indicates that discipline in attending school is a crucial factor in supporting academic achievement.

Previous scores have a positive effect on current exam scores. This indicates that past academic performance remains an important indicator of future academic performance. Meanwhile, tutoring sessions also have a positive effect, meaning that additional learning support outside the classroom can help students improve their exam results.

Regarding family factors, students with low parental involvement have average test scores nearly 2 points lower than students with high parental involvement. Similarly, regarding the learning environment, students with low access to learning resources have test scores approximately 2 points lower than those with high access. These findings confirm that differences in academic achievement stem not only from students' own study efforts but also from the social support and learning resources available to them.

IV. DISCUSSION

The research findings indicate that academic factors have the most significant impact on students' test scores. Attendance and study hours are the two variables that most consistently show a positive influence. This finding makes sense because students who attend class more frequently have a greater opportunity to follow the teacher's explanations, understand the material, and participate in the learning process. Similarly, students who spend more time

studying tend to have a better understanding of the material and are better prepared when taking exams.

The positive influence of previous scores indicates continuity in academic performance. Students with a strong academic foundation tend to maintain their performance in subsequent evaluations. This suggests that educational interventions should be implemented as early as possible, as learning achievements in earlier stages can influence outcomes in subsequent stages.

Findings regarding tutoring sessions show that additional academic support remains relevant in improving learning outcomes. Tutoring can serve as a means of reinforcing material, deepening concepts, and addressing students' learning weaknesses. Thus, academic support programs can be one strategy to improve student achievement, particularly for those with lower initial performance.

In the family dimension, parental involvement has been shown to significantly influence test scores. This involvement can be manifested through monitoring of study habits, providing motivation, or creating a conducive learning environment at home. High levels of parental support appear to reinforce students' study habits and enhance their academic responsibility.

In the learning environment dimension, access to learning resources also plays a crucial role. Students with high access to learning resources achieve better scores compared to those with limited access. This indicates that the quality of the learning environment including instructional materials, supporting tools, and learning facilities serves as a supportive factor that enhances the effectiveness of students' learning efforts.

Overall, the results of this study indicate that students' academic achievement is multidimensional. Test scores are determined not only by internal student factors but also by the quality of external support received. Therefore, efforts to improve learning outcomes should be undertaken in an integrated manner, involving students, families, and the learning environment.

V. CONCLUSIONS

This study concludes that academic, family, and learning environment factors simultaneously have a significant influence on students' exam scores. The regression model indicates that the six variables tested account for 66.0% of the variation in students' exam scores. Partially, study hours, attendance, previous scores, and tutoring sessions have a positive and significant effect on exam scores. Additionally, students with high parental involvement and high access to learning resources tend to achieve better exam results compared to students in the moderate or low categories.

These findings indicate that improving academic achievement cannot be focused solely on students as individuals. Family support and the availability of an adequate learning environment are also important

factors that must be considered. Thus, strategies to improve learning outcomes need to be implemented comprehensively through strengthening study discipline, improving attendance, providing tutoring, fostering parental involvement, and providing better learning resources.

REFERENCES

- [1] C. Ortiz-De-villate, J. Rodríguez-Santero, and J. J. Torres-Gordillo, "Contextual, personal and family factors in explaining academic achievement: A multilevel study," *Sustain.*, vol. 13, no. 20, 2021, doi: 10.3390/su132011297.
- [2] I. Martínez, "Higher education student engagement in learning activities: Clarifying concepts and introducing a short-scale," pp. 1–17, 2026, doi: 10.1371/journal.pone.0340391.
- [3] E. Tomul, E. Önder, and E. Taslidere, "The relative effect of student, family and school - related factors on math achievement by location of the school," *Large-scale Assessments Educ.*, pp. 1–29, 2021, doi: 10.1186/s40536-021-00117-1.
- [4] M. Latifah and R. Amelia, "Predictors of Adolescent Academic Achievement: The Role of Individual and Family Socioeconomic Factors," *J. Fam. Sci.*, vol. 4, no. 2, pp. 105–119, 2020, doi: 10.29244/jfs.4.2.105-119.
- [5] E. Caponera and B. Losito, "Context factors and student achievement in the IEA studies: evidence from TIMSS," *Large-scale Assessments Educ.*, 2016, doi: 10.1186/s40536-016-0030-6.
- [6] Y. S. Wibowo, F. A. Setiawati, S. R. Qodriah, E. Nizeyumukiza, and Y. Ayriza, "Do School Climate and Subjective Well-Being Affect Student Achievement in Indonesia?: A Linear Regression Analysis," *J. Pendidik. Progresif*, vol. 10, no. 2, pp. 183–191, 2020, doi: 10.23960/jpp.v10.i2.202004.
- [7] M. E. L. Jihaoui, "Factors Affecting Student Academic Performance: A Combined Factor Analysis of Mixed Data and Multiple Linear Regression Analysis," *IEEE Access*, vol. 13, no. December 2024, pp. 15946–15964, 2025, doi: 10.1109/ACCESS.2025.3532099.
- [8] R. S. J. D. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," vol. 1, no. 1, pp. 3–16, 2009.
- [9] M. U. Nisa, D. Areva, and V. Eprillison, "The Influence of Learning Styles, Learning Independence, Family Support, and School Environment on Academic Achievement with Learning Motivation as a Moderating Variable in 11th Grade / Fase F Economics Students," vol. 6, no. 1, pp. 662–673, 2026, doi: 10.58737/jpled.v6i1.807.
- [10] N. K. R. Kartika, N. Natajaya, and K. Rihendra, "Kualitas Pembelajaran Terhadap Prestasi Belajar Siswa Pada Mata Pelajaran Ekonomi," *e-Journal Progr. Pascasarj. Univ. Pendidik. Ganesha Progr. Stud. Adm. Pendidik.*, vol. 4, 2023.
- [11] Z. Zhao, P. Ren, and Q. Yang, "Student Self-Management, Academic Achievement: Exploring the Mediating Role of Self-Efficacy and the Moderating Influence of Gender—Insights From a Survey Conducted in 3 Universities in America," no. June 2023, pp. 1–12.
- [12] Y. Utami *et al.*, "Analisis Pengaruh Faktor Sosial terhadap Kinerja Akademik Mahasiswa Menggunakan Regresi Linier," vol. 24, pp. 1–8, 2025.
- [13] R. Saragih, I. Gunawan, and I. Parlina, "Implementasi Metode Regresi Linier Berganda untuk Prediksi Siswa Berprestasi Berdasarkan Status Sosial dan Kedisiplinan Implementation of the Multiple Linear Regression Method to Predict Student Achievement Based on Social Status and Discipline," vol. 2, no. 2, pp. 133–142, 2023, doi: 10.55123/jomlai.v2i2.3128.
- [14] O. El Aissaoui, Y. El, A. El, L. Oughdir, A. Dakkak, and Y. El Alloui, *A Multiple Linear Regression-Based Approach to Predict Student Performance*, vol. 1. Springer International Publishing, 2020. doi: 10.1007/978-3-030-36653-7.
- [15] B. K. Slinker and S. A. Glantz, "Statistical Primer for Cardiovascular Research Dependent Variable," pp. 1732–1737, 2008, doi: 10.1161/CIRCULATIONAHA.106.654376.
- [16] A. Mohamed, W. Husain, and A. Rashid, "The Third Information Systems International Conference A Review on Predicting Student's Performance using Data Mining Techniques," *Procedia - Procedia Comput. Sci.*, vol. 72, pp. 414–422, 2015, doi: 10.1016/j.procs.2015.12.157.