ICDSOS
The 3rd International Conference on Data Science and Official Statistics
November 27 - 28, 2025

"From Data to Impact: Leveraging Innovation in Data Science and Official Statistics for Data-Driven Policy Making and Global Competitiveness"

Deciphering Student Academic Success: Bayesian Analytical Insights

V Suriya[1,*], S Lakshmi[1], and Reshmavathi[1]

[1] PG Department of Biostatistics, SDNB Vaishnav College for Women (Autonomous), Chennai – 600044, India

*Corresponding author's email: suriya.v@sdnbvc.edu.in

Abstract. This study delves into the factors influencing student's academic achievement utilizing Bayesian mixed effect models. It presents five distinct models, each integrating various fixed variables such as gender, playing hours, stress level, and travelling hours, alongside random variables such as school level and type of school. These models are evaluated using the Leave-One-Out Information Criterion (LOOIC) to gauge their adequacy in fitting the data and predicting outcomes. The findings unveil that the inclusion of additional factors, such as school characteristics and students' activities, modifies the relationship between gender and academic success, with gender exerting a diminishing influence as more variables are incorporated. Additionally, stress level and travelling hours emerge as noteworthy predictors of average marks. Among the models assessed, the one incorporating gender, playing hours, and stress level as fixed effects, alongside school level and type as random effects, demonstrates superior fit and predictive capability. This underscores the significance of considering both individual traits and contextual elements in comprehending academic performance.
Keyword: Bayesian mixed effect model, gender-based disparities, fixed effects, random effect, Level-One-Out-Information Criterion (LOOIC).

1. Introduction

Education is fundamentally important as it equips individuals with the knowledge, skills, and opportunities necessary for personal growth and success. It empowers individuals to pursue their passions, make informed decisions, and contribute meaningfully to society. Additionally, education plays a crucial role in driving economic development, fostering social cohesion, and promoting political stability. Through education, individuals can access better employment opportunities, contribute to innovation and productivity, and engage in democratic processes effectively. Moreover, education is essential for addressing global challenges and achieving sustainable development goals, making it a cornerstone of progress and prosperity for individuals and societies alike.

Bayesian multilevel modelling, also known as hierarchical linear modelling (HLM) or mixed-effects modelling, is a powerful statistical technique used to analyze complex data structures with hierarchical or nested levels. It allows researchers to account for variability within and between different levels of analysis, such as individuals within groups or clusters. By incorporating both fixed and random effects, Bayesian multilevel modelling offers flexibility and precision in capturing the nuances of relationships within the data. This method works especially well in educational research because the data often exhibit hierarchical structures, such as students nested within classrooms or schools.

Through Bayesian multilevel modelling, researchers can uncover insights into the factors influencing educational outcomes and make informed decisions to improve educational practices and policies.In this study, data were collected through a structured questionnaire from 277 students representing a wide range of geographical regions, school types, and socioeconomic backgrounds. The sample includes students from rural, town, urban, and city areas, and from different school systems such as government, aided, and private institutions, as well as State Board, CBSE, and ICSE curricula. This diversity ensures that the dataset captures variations in educational access, learning environments, and contextual factors influencing student performance. By employing a Bayesian multilevel research method, the analysis appropriately accounts for the nested structure of the data includes students within schools and allows for the estimation of both individual-level and school-level effects on academic achievement. Such representation enhances the generalizability of the findings and provides a more comprehensive understanding of how demographic, institutional, and contextual variables interact to shape student academic success.

2. Research Method

Bayesian multilevel modelling is a robust statistical framework for analyzing hierarchical or nested data structures, commonly encountered in fields such as psychology, education, sociology, and epidemiology [7],[12]. Also referred to as hierarchical linear modelling (HLM) or mixed-effects modelling, it enables researchers to analyze complex relationships while accounting for variability at multiple levels of data.

Bayesian multilevel modelling has been increasingly applied across diverse research contexts, offering flexibility and interpretability in the study of hierarchical data. [15] Explored the use of Bayesian multilevel models through the *brms* software in linguistic research, focusing on gender effects on vowel variability in Standard Indonesian. Their work detailed hierarchical model implementation, prior specification, posterior interpretation, and model comparison using both *brms* and *lme4*, alongside Bayesian indices such as LOOIC and WAIC. In the field of education, [20] applied Bayesian multilevel modelling to investigate factors influencing eighth-grade students' academic performance in Ethiopia. Their findings revealed that student-level factors such as self-concept, study time, motivation, absenteeism, and family income, along with school-level factors like instructional strategies and teacher-student ratios, significantly affect achievement. The study also highlighted the gap between private and government schools, emphasizing the need for targeted strategies to reduce disparities in educational outcomes. From a methodological standpoint, [4] provided a comprehensive overview of Bayesian multilevel techniques using the *brms* R package, focusing on the theoretical foundation of Bayesian inference and demonstrating its application in complex hierarchical data analysis. Several other studies have also applied Bayesian statistical modelling approaches to educational and health-related data, providing additional context for this work [1]–[3], [5], [6], [8]–[11], [13], [14], [16]–[19]. Together, these studies demonstrate the versatility and strength of Bayesian multilevel modelling in addressing nested data structures, making it a valuable tool for applied research.

At the core of this framework lies the Bayesian principle of treating model parameters as random variables. Unlike classical (frequentist) approaches that treat parameters as fixed but unknown, Bayesian methods incorporate uncertainty directly into parameter estimation by combining prior knowledge with observed data. This concept is formalized through Bayes' theorem:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \quad (1)$$

In this equation, $P(\theta|y)$ represents the posterior distribution, reflects updated beliefs about parameters after observing the data. $P(y|\theta)$ is the likelihood function, describing the probability of the observed data given specific parameter values. $P(\theta)$ denotes the prior distribution, representing existing knowledge before data collection, and $P(y)$ is the marginal likelihood ensuring the posterior is a valid probability distribution. The marginal likelihood also plays an essential role in model comparison using criteria such as WAIC and LOOIC, which assess predictive performance.

ICDSOS
The 3rd International Conference on Data Science and Official Statistics
November 27 - 28, 2025

A major strength of Bayesian multilevel modelling is its ability to estimate fixed and random effects simultaneously. Fixed effects represent parameters that remain constant across all groups, capturing the average influence of predictors, while random effects account for unobserved heterogeneity across groups, such as differences in school environments or teaching quality that may influence student outcomes.

The present study employed Bayesian multilevel modelling to predict students' average marks based on multiple explanatory variables. This approach was chosen due to the hierarchical structure of the data, where students are nested within schools that differ in academic resources, environment, and instructional quality. Traditional regression models often overlook such clustering, potentially leading to biased parameter estimates and underestimated uncertainty. In contrast, Bayesian multilevel modelling partitions variability across multiple levels, offering more accurate and reliable inferences about both individual- and group-level effects.

The model incorporated gender, playing hours, stress level, and travelling hours as fixed effects with their respective reference categories, while level of school and type of school were included as random effects to capture unobserved variability across institutions. When all predictor variables are set at their reference levels, the model estimates the intercept value ($\propto$), representing the expected average marks for a baseline student profile.

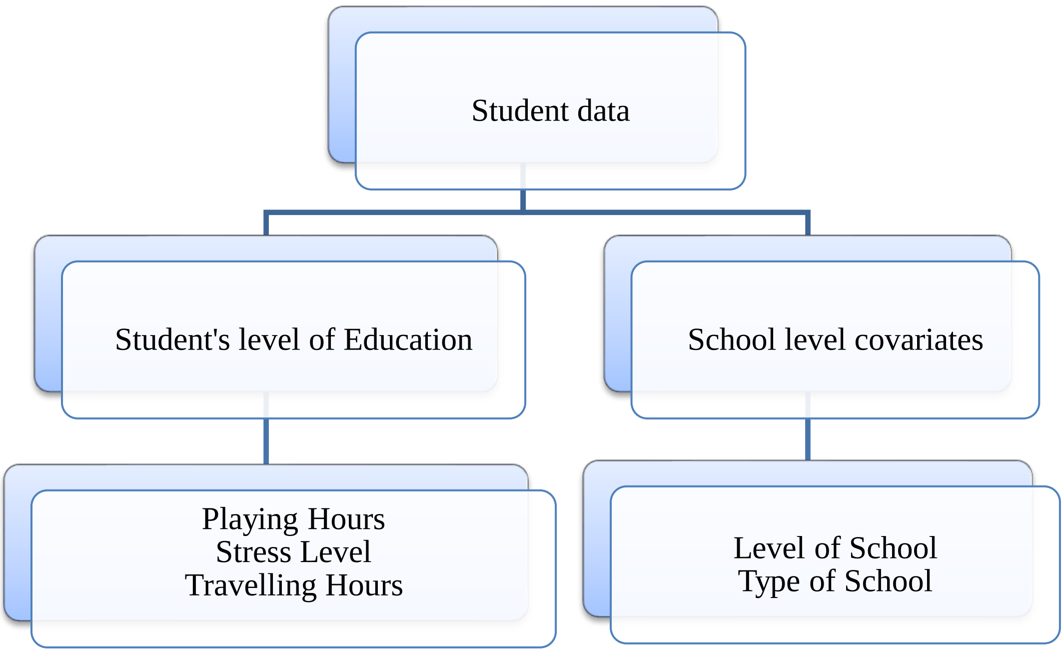

Figure 1. Data Hierarchy in Bayesian Multilevel Modelling

In this study, the Bayesian framework provided a coherent probabilistic structure for estimating both fixed and random effects while quantifying uncertainty around them. The parameter estimates for all models are presented in the following table, including their corresponding mean values, standard errors (SE), 95% credible intervals, and the Gelman–Rubin statistic ($\hat{R}$), a diagnostic measure used to assess model convergence in Bayesian estimation.

3. Results And Discussion

This section presents the results of the Bayesian multilevel modelling carried out to identify the factors influencing students' academic performance. A series of five models were developed, each progressively incorporating additional predictors and random effects to better explain the variation in average marks among students. The fixed effects in these models include gender, playing hours, stress level, and travelling hours, while the random effects represent differences across school levels and school types. Model comparison was performed using the Leave-One-Out Information Criterion (LOOIC) to evaluate the goodness of fit and predictive accuracy of each model. The results of each model are described in detail below.

3.1. Model 1

Model 1 presents predicted average marks for students, considering gender as a fixed effect, with males serving as the reference category.

Table 1. Prediction of Student Average Marks with Gender as a Fixed Effect

Parameters	Estimate	SE	Lower bound	Upper bound	Rhat
$\propto$	63.88	1.24	61.41	66.35	1.00
$\boldsymbol{\beta_1}$**(Gender)**					
Female	6.65	1.65	3.37	9.86	1.00
$\boldsymbol{\sigma_e}$	13.83	0.60	12.73	15.07	1.00

The intercept ($\propto$) of 63.88 represents the estimated baseline value of the outcome variable when all predictor variables are at their reference levels (if any). The coefficient for the predictor variable Gender (Female) is estimated to be 6.65. This suggests that, on average, the outcome variable is expected to be 6.65 units higher for females compared to males.

3.2. Model 2

Model 2 estimates the expected average marks for student by taking level of school and type of school as random effects and gender as a fixed effect, with males acting as the reference group.

Table 2. Prediction of Student Average Marks with Gender as Fixed Effect and School Level and Type as Random Effects

Parameters	Estimate	SE	Lower bound	Upper bound	Rhat
$\propto$	66.07	5.14	56.03	77.31	1.00
$\boldsymbol{\beta_1}$**(Gender)**					
Female	5.51	1.79	1.98	9.02	1.00
$\boldsymbol{\sigma_{level\ of\ school}}$	7.28	5.86	0.41	22.68	1.00
$\boldsymbol{\sigma_{Type\ of\ scchool}}$	2.71	3.22	0.07	11.68	1.00
$\boldsymbol{\sigma_e}$	13.73	0.59	12.64	14.96	1.00

In model 2, the estimated intercept increases from 63.88 to 66.07, while the coefficient for Gender (Female) decreases from 6.65 to 5.51 compared to model 1. This suggests that including additional variables like school level and type alters the relationship between gender and average marks, reducing the estimated impact of gender on academic performance. The standard deviation of average marks across different school levels is estimated at 7.28, indicating greater variability, while the standard deviation across different types of schools is 2.71, suggesting significant influence of school type on academic performance.

3.3. Model 3

Model 3 shows the expected average marks for each student by taking level of school and type of school as random effects. Gender and playing hours are fixed effects, with male and non-playing students as reference categories.

Table 3. Prediction ofStudent Average Marks with Gender and Playing Hours as Fixed Effects and School Level and Type as Random Effects

Parameters	Estimate	SE	Lower bound	Upper bound	Rhat
$\propto$	69.83	5.74	58.80	82.84	1.00
$\boldsymbol{\beta_1}$ **(Gender)**					
Female	4.71	1.80	1.20	8.25	1.00
$\boldsymbol{\beta_2}$ **(Playing hours)**					
< 1 hour	-2.16	2.55	-7.22	2.79	1.00
1 hour	-3.06	2.27	-7.47	1.38	1.00
> 1 hour	-5.64	2.40	-10.40	-0.99	1.00
$\boldsymbol{\sigma_{level\ of\ school}}$	8.35	6.57	0.42	25.49	1.00
$\boldsymbol{\sigma_{Type\ of\ scchool}}$	2.68	3.46	0.07	12.16	1.00
$\boldsymbol{\sigma_e}$	13.64	0.60	12.53	14.86	1.00

When playing hours are considered as a fixed effect, the baseline average mark increases to 69.83 and the gender coefficient decreases to 4.71, indicating a further reduction in the impact of gender on average marks when school-related factors are accounted for. Students playing less than an hour are expected to score 2.16 points lower than non-players, while those playing for an hour are expected to score 3.06 points lower, and those playing more than an hour are expected to score 5.64 points lower. The school level's standard deviation is estimated at 8.35, suggesting increased variability with the inclusion of playing hours. However, the predicted standard deviation for school type decreases to 2.68, indicating variations in playing hours across different grade levels, possibly due to higher focus on studies among high school students.

3.4. Model 4

Model 4 resembles Model 3, but now includes stress level as an additional fixed factor, with the absence of stress as the reference category.

Table 4. Predictionof Student Average Marks with Gender, Playing Hours, and Stress Level as Fixed Effects and School Level and Type as Random Effects

Parameter	Estimate	SE	Lower bound	Upper bound	Rhat
$\propto$	73.58	6.13	61.79	86.81	1.00
$\boldsymbol{\beta_1}$**(Gender)**					
Female	4.53	1.80	0.98	8.07	1.00
$\boldsymbol{\beta_2}$ **(Playing hours)**					
< 1 hour	-1.91	2.51	-6.85	3.07	1.00
1 hour	-3.56	2.27	-8.03	0.89	1.00
> 1 hour	-6.05	2.39	-10.75	-1.41	1.00
$\boldsymbol{\beta_3}$ **(stress level)**					
Sometimes	-3.10	2.46	-7.91	1.74	1.00
Often, always	-6.42	2.76	-11.80	-1.02	1.00
$\boldsymbol{\sigma_{level\ of\ school}}$	8.22	5.92	1.20	24.00	1.00
$\boldsymbol{\sigma_{Type\ of\ scchool}}$	2.82	3.46	0.09	12.18	1.00
$\boldsymbol{\sigma_e}$	13.55	0.59	12.46	14.77	1.00

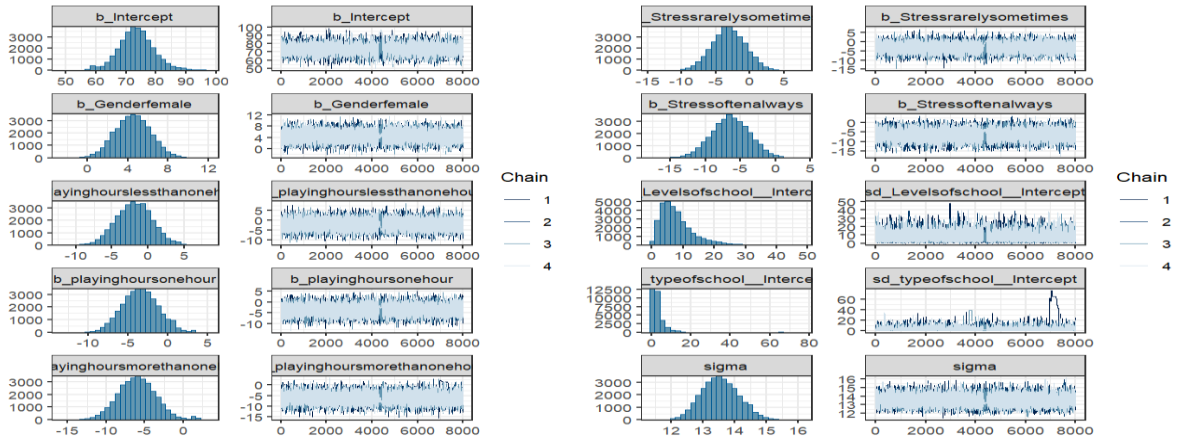

Figure 2. Histograms of posterior samples and trace plots of the intercept

This model has the highest intercept among three models, at 73.58. The coefficient for gender (female) consistently decreases across all models, dropping to 4.53 in this model, indicating a reduced impact of gender on average marks as more variables are added. For playing hours, the estimated impact remains consistent across models: 1.91 for less than an hour, 3.56 for an hour, and 6.05 for more than an hour of play. Higher stress levels are associated with lower average marks, with students experiencing occasional stress expected to score 3.10 points lower, and those experiencing consistent stress expected to score 6.42 points lower. Considering stress level, the standard deviation for school type increases to 2.82, while for school level it decreases to 8.22, suggesting stress affects school type more, possibly due to heavier workloads in private schools.

3.5. *Model 5*

Model 5 is same as model 4, but now includes travelling hours as an additional fixed factor, with travel times less than one hour serving as the reference category.

Table 5. Prediction of Student Average Marks with Gender, Playing Hours, Stress Level, and Travelling Hours as Fixed Effects and School Level and Type as Random Effects

Parameter	Estimate	SE	Lower bound	Upper bound	Rhat
$\propto$	71.89	6.13	59.6	84.49	1
β_1(Gender)					
Female	4.38	1.83	0.75	7.92	1
β_2(Playing hours)					
< 1 hour	-1.47	2.56	-6.54	3.53	1
1 hour	-3.27	2.32	-7.77	1.31	1
> 1 hour	-5.33	2.43	-10.18	-0.64	1
β_3(Stress level)					
Sometimes	-3.17	2.45	-8	1.66	1
Often	-6.56	2.75	-12	-1.17	1
β_4(Travelling hours)					
1 hour	2.73	1.97	-1.08	6.64	1
2 hours	2.71	2.35	-1.87	7.24	1
3 hours	-0.32	5.84	-11.79	11.16	1
$\sigma_{Level\ of\ school}$	8.47	6.08	1.35	24.76	1
$\sigma_{Type\ of\ school}$	2.65	3.58	0.07	12.49	1

σ_e	13.56	0.59	12.47	14.77	1

In Model 5, the baseline average mark slightly decreases to 71.89 compared to Model 4, indicating a small effect of traveling hours on academic performance. The coefficient for Gender (Female) also decreases slightly. Comparing stress level coefficients, Model 5 indicates a slightly higher influence on average marks than Model 4. Students traveling for one or two hours are predicted to have better average marks than those traveling for less than an hour, while those traveling for three hours may have 0.32 units lower mark than students travelling for less than an hour. The standard deviation for school level increases to 8.47, while for school type it drops to 2.65, this means travelling hours affect school level more than school type.

3.6. *Model comparison using LOOIC*

The table 4.6 represents a comparison of the five Bayesian models using the Leave-One-Out Information Criterion (LOOIC), a measure of model fit and predictive accuracy. Lower LOOIC values indicate better predictive performance.

Table 6. Comparison of Bayesian Models Using Leave-One-Out Information Criterion

Model	LOOIC	SE	Δ LOOIC	Δ SE	Formula
bmod5	2243.9	21.8	0.00	1.6	Gender+ Playing hours + Stress Level+ Travelling hours+(1| Level of school) + (1|Type of school)
bmod4	2240.0	21.5	3.9	0.0	Gender+ Playing hours + Stress Level+(1| Level of school) + (1|Type of school)
bmod3	2242.0	20.9	1.9	2.5	Gender+ Playing hours+(1| Level of school) + (1|Type of school)
bmod2	2242.1	20.8	1.8	3.1	Gender+(1| Level of school) + (1|Type of school)
bmod1	2243.4	20.6	0.5	3.8	Gender

Among the candidate models, Model 4 has the lowest LOOIC value (2240.0), suggesting it provides the best balance between explanatory power and predictive accuracy. Although Model 5 includes an additional predictor (Travelling Hours), its LOOIC (2243.9) is slightly higher than Model 4, indicating that the extra variable does not substantially improve out-of-sample prediction. Therefore, Model 4 incorporating Gender, Playing Hours, and Stress Level as fixed effects, along with Level of School and Type of School as random effects emerges as the preferred model for predicting student average marks.

From this adding more predictors does not always enhance predictive accuracy. In this case, travelling hours may not provide substantial independent explanatory power beyond the effects already captured by other predictors such as stress level and playing hours. Travelling hours may also be correlated with other variables or with the random effects of level of school and type of school, as these factors already account for contextual differences such as distance, accessibility, and school environment. As a result, the inclusion of Travelling hours increases model complexity without a meaningful reduction in prediction error. The LOOIC penalizes unnecessary complexity and rewards models that generalize well to new data, showing that Model 4 achieves a better balance between fit and simplicity. While Model 5 may fit the observed data slightly better, Model 4 generalizes more effectively by avoiding over fitting and capturing the essential structure of the data with fewer parameters. This finding emphasizes the value of model simplicity and demonstrates that increased complexity does not always lead to better predictive performance in Bayesian multilevel modelling.

3.7. Discussion

The present study employed Bayesian multilevel modelling to examine the influence of individual and school-level factors on student's academic performance. Through the integration of fixed and random effects, the analysis captured variations among students across different types and levels of schools. The results provide a detailed understanding of the multi factorial determinants of academic achievement.

Gender appeared as a significant predictor in all models, with female students consistently attaining higher average marks than male students. The gender difference became less pronounced in extended models, indicating that part of the variation is explained by contextual and behavioural factors such as stress level and playing hours. This outcome supports earlier studies suggesting that differences in learning behaviour, study consistency, and stress management contribute more to academic disparities than inherent ability.

Playing hours showed a negative association with academic performance, particularly as they extended beyond one hour per day. This observation aligns with prior evidence that excessive engagement in non-academic activities reduces focus on study time. Moderate recreational activity may contribute to mental well-being, though the effect in this dataset was not statistically strong.

Stress level proved to be one of the most influential factors, with higher stress associated with reduced academic achievement. Students reporting frequent stress displayed noticeably lower marks than those experiencing minimal stress. This finding is consistent with educational and psychological research indicating that persistent stress limits concentration, memory, and motivation, which are essential for academic progress. Including stress level in the model significantly improved the overall model fit, confirming its central role in determining performance.

Travelling hours exerted a modest yet complex influence. Students commuting for one to two hours tended to perform slightly better than those travelling shorter distances, possibly reflecting access to schools with superior educational resources in urban areas. However, travel exceeding three hours was associated with lower marks, likely due to tiredness and limited study time. These results are in agreement with previous research noting that travel duration can function as both a barrier and an opportunity depending on school accessibility and transportation conditions.

Random effects for school level and type demonstrated clear variability in student performance across institutions. This variability highlights the role of institutional factors such as teaching quality, learning environment, and resource availability in shaping academic outcomes. The Bayesian framework enabled reliable estimation of these differences while accounting for data hierarchy and uncertainty, producing a realistic representation of the educational structure.

Model comparison using the Leave-One-Out Information Criterion (LOOIC) indicated that the model incorporating gender, playing hours, and stress level as fixed effects and school level and type as random effects achieved the strongest predictive accuracy. The inclusion of travelling hours did not yield additional improvement, illustrating the need for a balanced model that captures essential determinants without excessive complexity. The selected model therefore provides an optimal balance between precision and interpretability, reflecting both personal and institutional dimensions of academic performance.

4. Conclusion

This study highlights the effectiveness of Bayesian multilevel modelling in analyzing hierarchical educational data and identifying critical determinants of student academic success. The findings reveal that gender differences in performance diminish when contextual factors such as stress and extracurricular engagement are accounted for. Stress levels and excessive playing time are the most influential predictors of reduced academic achievement, while moderate travel distances may have a neutral or slightly positive effect.

The superiority of the model incorporating gender, playing hours, and stress level underscores the importance of addressing both behavioural and psychological determinants in educational planning. Interventions aimed at reducing student stress, promoting balanced leisure activities, and enhancing institutional support systems could substantially improve academic outcomes. Furthermore, the random

ICDSOS
The 3rd International Conference on Data Science and Official Statistics
November 27 - 28, 2025

effects analysis demonstrates that school level factors continue to play a meaningful role, emphasizing the need for equitable resource distribution and supportive learning environments.

Overall, the Bayesian framework provides a flexible, probabilistic approach to understanding educational phenomena, accommodating uncertainty and complexity more effectively than traditional frequentist methods. Future research could extend this analysis by incorporating longitudinal data, exploring subject specific performance, or integrating additional socio-economic variables to further enhance the predictive understanding of academic success.

References

[1] S. P. Baliyan and D. Khama, "How distance to school and study hours after school influence students' performance in mathematics and English: A comparative analysis," *J. Educ. e-Learning Res.*, vol. 7, no. 2, pp. 209–217, 2020.

[2] B. Balabanian, *Transportation and Its Effect on Student Performance*, M.S. thesis, Dept. Geogr. Inf. Sci., 2020.

[3] R. Bertolini, S. J. Finch, and R. H. Nehm, "An application of Bayesian inference to examine student retention and attrition in the STEM classroom," *Front. Educ.*, vol. 8, Art. no. 1073829, 2023, doi: 10.3389/feduc.2023.1073829.

[4] P. C. Bürkner, "brms: An R package for Bayesian multilevel models using Stan," *J. Stat. Softw.*, vol. 80, pp. 1–28, 2017.

[5] C. O. Duze, "Average distance travelled to school by primary and secondary school students in Nigeria and its effect on attendance," *Afr. Res. Rev.*, vol. 4, no. 4, 2010.

[6] A. Gelman, "Comment: Bayesian checking of the second levels of hierarchical models," *Stat. Sci.*, vol. 22, no. 3, pp. 349–352, 2007.

[7] A. Gelman, J. Hill, and A. Vehtari, *Regression and Other Stories*. Cambridge, U.K.: Cambridge Univ. Press, 2020.

[8] A. Gelman, A. Jakulin, M. G. Pittau, and Y.-S. Su, "A weakly informative default prior distribution for logistic and other regression models," *Ann. Appl. Stat.*, vol. 2, no. 4, pp. 1360–1383, 2008, doi: 10.1214/08-AOAS191.

[9] A. Gelman, A. Vehtari, D. Simpson, C. C. Margossian, B. Carpenter, M. Yao, and M. Modrák, "Bayesian workflow," *arXiv preprint arXiv:2011.01808*, 2020, doi: 10.48550/arXiv.2011.01808.

[10] J. Huang, K. Yang, Q. Wang, P. Yang, Z. Ruan, J. Wang, and Z.-Q. Zhang, "Bayesian deep multi-instance learning for student performance prediction based on campus big data," *Neurocomputing*, vol. 647, Art. no. 130538, 2025, doi: 10.1016/j.neucom.2025.130538.

[11] J. K. Kruschke, *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. London, U.K.: Academic Press, 2015.

[12] R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton, FL, USA: CRC Press, 2020.

[13] J. A. Mhiliwa, *The Effects of School Distance on Students' Academic Performance: A Case of Community Secondary Schools in Makambako Town Council*, Ph.D. dissertation, Open Univ. of Tanzania, 2015.

[14] M. Mosia, F. O. Egara, F. A. Nannim, and M. Basitere, "Bayesian hierarchical modelling of student academic performance: The impact of mathematics competency, institutional context, and temporal variability," *Educ. Sci.*, vol. 15, no. 2, p. 177, 2025, doi: 10.3390/educsci15020177.

[15] L. Nalborczyk, C. Batailler, H. Lœvenbruck, A. Vilain, and P. C. Bürkner, "An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian," *J. Speech Lang. Hear. Res.*, vol. 62, no. 5, pp. 1225–1242, 2019.

[16] R. L. Peach, S. F. Greenbury, I. G. Johnston, S. N. Yaliraki, D. J. Lefevre, and M. Barahona, "Understanding learner behaviour in online courses with Bayesian modelling and time series characterisation," *Sci. Rep.*, vol. 11, no. 1, p. 2823, Feb. 2021, doi: 10.1038/s41598-021-81709-3.

[17] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York, NY, USA: Springer, 2004.

[18] S. Begum and S. S. Padmannavar, "Student performance analysis using Bayesian optimized random forest classifier and KNN," *Int. J. Eng. Trends Technol.*, vol. 71, no. 5, pp. 132–140, 2023, doi: 10.14445/22315381/IJETT-V71I5P213.

[19] S. Salleh and Y. Yassin, "Bayesian model for academic performance prediction in learning analytics," *Int. J. Acad. Res. Prog. Educ. Develop.*, vol. 13, 2024, doi: 10.6007/IJARPED/v13-i1/20364.

[20] N. Y. Sebro and A. T. Goshu, "Modelling of academic achievement of primary school students in Ethiopia using Bayesian multilevel approach," *J. Educ. Learn.*, vol. 6, no. 1, pp. 337–346, 2017.

ICDSOS
The 3rd International Conference on Data Science and Official Statistics
November 27 - 28, 2025