The Lighthouse Journal of Computational & Numerical Sciences, Vol. 01, Issue 01, Winter 2022, pp. 13-32

Multilevel Modeling Approach for Assessing the Performance of Primary School Students

Qamruz Zaman[1], Muhammad Waqas[2], Abdurrahman Sabir[1], Himayat Ullah[3]

ABSTRACT

Education is extremely important in today's society. The primary goal of education, as well as educational institutions, is to improve a person's physical, mental, and social well-being. The current study not only focuses on the impact of basic education, but also investigates the factors that influence students' and institutes' academic performance at various levels using a statistical multilevel model. The purpose of this study was to determine the academic performance of primary school pupils (especially in the fourth and fifth grades) and their institutions in Peshawar. The study's goals were to examine the performance of primary school children in both private and public schools, as well as their gender and socioeconomic background. The data was acquired using a cross-sectional data collection approach from 346 schools with a total of 2565 kids, 1259 of whom attended government primary schools and 1305 of whom attended private primary schools. Female students made up 49.7% of the sample, while male students made up 50.3 percent. The dependent variable is the performance of elementary school students, which is measured by their exam scores. The results of the study revealed that gender had a substantial impact on outcomes, but parental support had no bearing. The importance of parents' education and the availability of tuition played a significant effect.

Keywords: *Statistics, Multilevel Modelling, Variance, Gender, Students*

INTRODUCTION

Modelling, whether in Mathematics or Statistics, is one of the most significant tools in any analytical investigation. It is, in fact, a representation of a real-life dilemma. The idea is most commonly used in statistics for making inferences. Different statistical tools/approaches are routinely employed for data modelling, with the regression phenomenon being one of the most popular. Under normality assumptions, the technique

[1] Department of Statistics, University of Peshawar, Pakistan, **Corresponding Author's Email: qamruzzaman@uop.edu.pk**
[2] Department of Statistics, University of Swabi, Pakistan
[3] Hayatabad Medical Complex, Peshawar, Pakistan

Manuscript Received: May 13, 2022
Manuscript Accepted: Dec 16, 2022

is based on the relationship between continuous dependent and independent variables (Draper & Smith 1998).

Statistical analysis is not a novel instrument. It was originally utilized in the 18th century, and researchers have made substantial advances to the discipline since then. There are numerous limits to regression modeling, which scholars in this discipline have worked hard to address (Goldstein 1989). The characteristics of a child's classroom, school, and public education system significantly affect how well they are doing academically (Luke 2019, Televantou 2015).

Common regression methods do not give any attention to the hierarchical relation. However, this problem can be solved by using the generalized form of regression analysis i.e., multilevel modeling (MLM). The method is consisting of not only all the characteristics of conventional regression but it covers the aspect of data reduction too (Raudenbush & Bryk 1986; Snijders & Bosker 1999; Kreft & De-Leeuw 1998).

Multilevel models are divided into two steps. Step1 is related to regression model and second phase emphasize on covariance. The first is similar to multiple regressions and the second is used for time related data which may be longitudinal and repeated data (Cohen & Cohen 1998; DiPrete & Grusky 1990). MLM is commonly used in every field of research including numerical, biological, applied fields etc. (Goldstein & Spiegelhalter 1996). It is commonly used for developing and exploring the relationship at different levels i.e., suitable for hierarchical relationship. The assumption of in-dependency does not play any important role in MLM unlike the conventional regression modeling (Goldstein 2003). One of the advantages of MLM is that it handles the structuring of the data, which is not possible in conventional methods (Huttner & Vanden 1993).

The area is still under developed as compared to the other statistical techniques but its importance cannot be ignored even in the field of ecological (Robinson 1950). Role of MLM is not restricted only to the estimation by conventional method, but is also expanded to the Bayesian estimation and expectation maximization (EM) Algorithm (Lindley & Smith 1972; Dempster et al. 1977). The development of aggregation analysis opened a new direction in the field of MLM (Boyd & Iversen 1979). Bryan and Jenkins (Tate & Hokanson 1993) used the MLM techniques to measure the effects of countries on individual. Their study revealed that the reliable results are obtained by using the hierarchical models. The fields of psychology and behaviour are also not free from MLM. In these, the outcome of interest is the result of the combination of different levels. The study of these factors is not possible without considering the hierarchical technique of statistics (De Leeuw & Kref 1985).

AIMS & OBJECTIVES

The following are the aims and objectives of the study:

1. To study the academic performance of private sector pre-school students using hierarchical structure of data.

2. To compare academic performance Pre-schools students in private sectors.

3. To compare the academic performance gender wise at school level.

4. Compare the academic performance of students with socioeconomic status

Methodology

Multilevel Modeling

With the beginning of 1980, multilevel modeling was introduced while studying the essence of science in approaches to education, but it quickly found applications in other disciplines and academic fields. Developing statistical software enhances progress in multi-level modeling application. Multi-level model is a statistical model that analyzes data in a single model from different hierarchical structure or level (Berkhof & Snijders 2002). It is an extension of the framework of multiple linear regression and traditional numerical multivariate techniques. In first attempt, the multi-level model estimates the fixed effect for all covariates of model and random effect for higher order covariates by analyzing variance covariance of the data. Instead of summarizing individual variables, these high order covariates are characterized by few distribution parameters. In the next step, the methodology of Bayes is used to derive the higher-order residual covariates of the Darling et al. model (Darling et al. 2005).

Two-level model

Two linear models are described and estimated at the same time in two-level MLM. The first model establishes relationships between lower-level units, while the second model aids in determining these links between lower-level units that fluctuate between higher-level units 1. The function at level 1 must satisfy the form below if the model only has one predictor variable.

$$\text{Level-1 } Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij}$$

where Y_{ij} is the response variable for i^{th} individual in the j^{th} group, X_{ij} is the explanatory variable related to i^{th} individual in the j^{th} group, β_{0j} is the intercept and j^{th} is the slope of j^{th} group and ε_{ij} is the error term of the i^{th} individual in the j^{th} group. It is assumed that the error term follows normal distribution with zero mean and unit variance (Raudenbush & Bryk 2002). The level-2 method analyzes as dependent variables by using the intercepts and slopes of the level-1 model. If the difference between groups is determined by one group level parameter, the level-2 model takes the form of an equation Level-2.

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}$$

where W_j is a group level variable, γ_{00} , γ_{01}, γ_{10} and γ_{11} are the regression coefficients u_{0j} and u_{1j} are the residuals. One typical assumes that for each group j, the random vector [u_{0j}, u_{1j}] follows multivariate normal distribution with each zero mean and variance of u_{0j} and $\text{var}\left(u_{ij}\right) = \delta_{ij}$, the covariance of $Cov\left(u_{ij}, u_{kj}\right) = \delta_{ii}$

Methods

This section offers a statistical study of primary data pertaining to primary school pupils' performance. The organized questionnaire collected information from primary school children in the Peshawar district. A total of 2565 primary school pupils were chosen from District Peshawar primary schools for this study. Different statistical tools, including descriptive and inferential statistics, are used in the statistical study. Inferential statistics contain multilevel modeling for discovering the significant elements that explain the performance of primary school kids, while descriptive statistics offer tabular presentation of data.

Sample Design

The representative sample was chosen using a stratified two-stage sampling approach. In the first stage, the number of schools was chosen at random, followed by the number of students from each strata i.e. Private and Government schools. The students from the schools were selected using simple random sampling procedure, where their attendance sheets were used as sampling frame. There were a total of 2430 schools, with 1039 government schools and 1391 private schools (Hartley & Rao 1967). There were 276040 pupils enrolled in public schools and 285874 students enrolled in private schools, respectively.

Sample Size Estimation

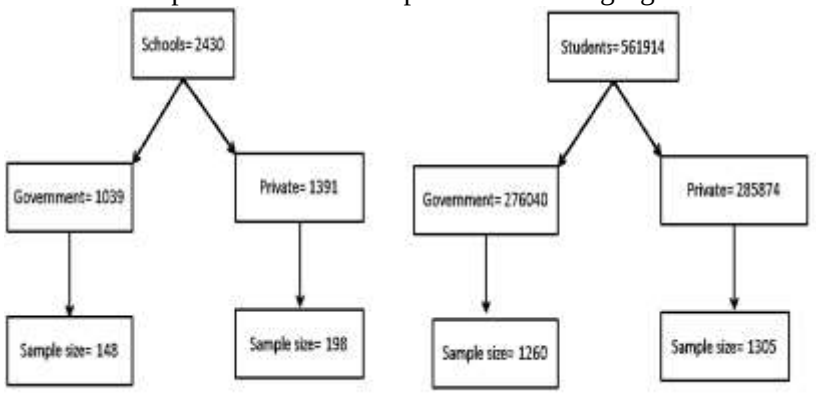
For the sample size calculation, (Krejice and Morgan 1970) formula is used.

$$n = \frac{\chi^2 * N * P * (1 - P)}{\left(ME^2 * (N - 1)\right) + (\chi^2 * P * (1 - P))} .$$

Where
n = Required Sample Size
χ^2 = Chi-Square statistic value for 1 d.f.
N = Population Size
P = 0.5 Population Proportion
ME = Margin of Error (expressed in proportion)
Out of 2430 schools, 346 are selected by considering 95% confidence interval and 5% margin of error. From the selected school's 2565 students are selected by using 1% margin of error.
Proportional allocation formula is used for selecting the representative sample.

$$n_i = \frac{n \times N_i}{N},$$

where
n= Sample size
N=Population size
n_i = Stratum sample size
N_i = Stratum size
The procedure can be explained with the help of the following figure.



Multilevel Modeling

Multilevel modeling approach was used for the analysis of data with the help of MLwiN package. Different selection procedures were used for selecting the best model like Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and likelihood ratio test.

Students' marks were considered as the response variable and age of the student, family type (nuclear, youth), size of the family, family income, school area and mode of education as independent variables. Suppose *i* denote students score and *j* the school, thenY_{ij} can be modeled as

$$Y_{ij} = \beta_{00} + \sum_{h=1}^{6} \beta_{hj} \; X_{hj} + \varepsilon_{ij} \, , \qquad (3.1)$$

Where β_{hj} are random regression coefficients and ε_{ij} is the error term?
Further, for the model at group or school level, β_{hj}depends on school type (Government Vs Private school), strength of the school and number of students in the class.
Then

$$\beta_{hj} = \gamma_{h0} + \sum_{g=1}^{3} \gamma_{hg} Z_{gh} + \quad \omega_{hj}. \qquad (3.2)$$

Whereβ_{hj}explain the dependency of j[th] factor (age, gender, family size, family type, income, area and mother education) on g^{th} factor (school type, school size and class size).

Therefore, the overall model is:

$$Y_{ij} = \beta_{00} + \sum_{h=1}^{6} (\gamma_{h0} + \sum_{g=1}^{3} \gamma_{hg} Z_{gh} + \omega_{hj}) X_{hj} + \varepsilon_{ij} \qquad (3.3)$$

Results

The statistical analysis of primary data on factors that influence pupils' performance at the primary level in Peshawar schools. The information was gathered from primary school kids. Multilevel modeling was used to examine the data. Model selection criteria were utilized to determine the best model utilizing the forward selection method of model selection. The structured questionnaire collected data from male and female pupils from both government and private sector schools in the district of Peshawar. A total of 346 schools, both private and public, were studied, with a total of 2565 students drawn from Peshawar's various regions. Different statistical tools, including descriptive and inferential statistics, are used in the statistical study. The Chi-Square test of independence, model diagnostic tests, Likelihood Ratio Test (LRT), Akaike Information Criteria (AIC), and Bayesian Information Criteria are examples of descriptive statistics, while inferential statistics include the Chi-Square test of independency, model diagnostic tests, Likelihood Ratio Test (LRT), Akaike Information Criteria (AIC), and Bayesian Information Criteria are examples of inferential statistics (BIC).

Descriptive Statistics

The following table 1 indicates the frequency distribution of the students of the Primary Schools at Level-I and the table 2 defines the frequency table of schools of district Peshawar at level-II.

Table 1: "Status of Student"

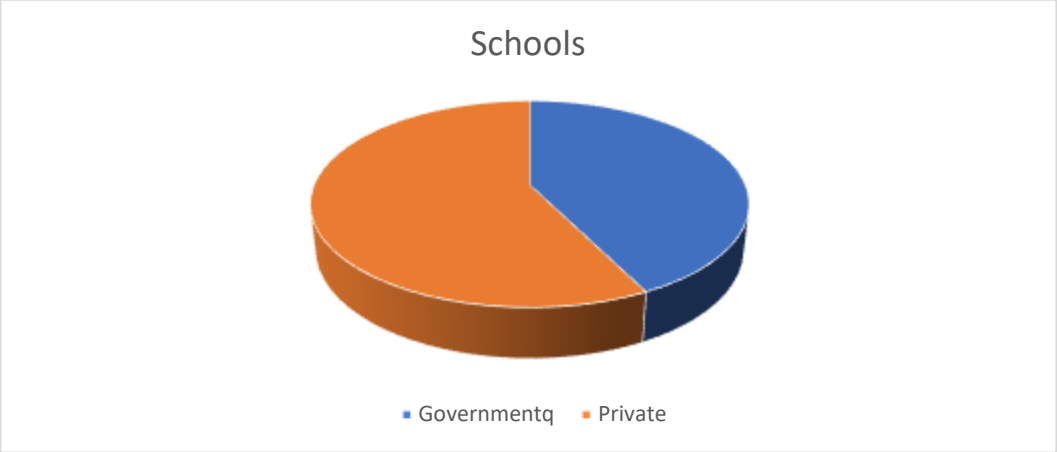
	Frequency	Percent
Government	1259	49.1
private	1305	50.9
Total	2564	100.0

The table defines that out of sample of 2564 students, 1259 students belonged to government primary schools at level-I with a percentage of 49.1 while 1305 students belonged to private primary schools at level-I with the percentage of 50.9. The following table 2 defines the frequency table of schools at district Peshawar.

Table 2: "School Type at District Peshawar"

	Frequency	Percent
Government	147	42.6
Private	198	57.4
Total	345	100.0

The above table defines that the data was collected from 147 government and 198 private sector schools at level-II with a percentage of 42.6 and 57.4 respectively. The following figure 1 indicates sector diagram of the status of the school of students at level-II.



The following table 4.3 defines the gender of the students at level-I. The defines that at level-I, there were 1275 male students and 1289 female students with the percentage of 49.7 and 50.3 respectively from the primary schools of district Peshawar. At level-II, there were 50 male and 50 female students from the Primary schools.

Table 3: "Gender of Student at Level-I"

	Frequency	Percent
male	1275	49.7
female	1289	50.3
Total	2564	100.0

The following table 4 defines the structure of family of the students at primary schools at level-I. From table 3, it is clear that 1264 primary level students belong to nuclear family structure with the percentage of 49.3 while 1300 students were from the joint family structure at primary school from various parts of district Peshawar.

Table 4: "Frequency Distribution of Family Type"

	Frequency	Percent
Nuclear	1264	49.3
Joint	1300	50.7
Total	2564	100.0

Comparison of Students Characteristics

The following table 5 presents the comparison of various characteristics of the students at primary school. It includes comparison of student's score in the class by the

various characteristics like gender of students, parent support, availing tuition facility and students in the academic. The table defines the mean and standard deviation of the performance (i.e. score of the student in class). The table compares the score of the student in class across various categories of the mentioned variable using student t-test. The table also include the p-value for statistical significance.

Table 5: "Comparison of Students Performance Using T-test**"**

Grouping Variable	Categories	Performance of student in class		t-Statistic	P-Value
		Mean	Standard Dev		
Gender	Male	326.5	15.26	14.365	0.000
	Female	389.6	18.89		
Parents Support	Yes	388.9	67.17	0.64	0.5278
	No	390.2	63.51		
Tuition facility	Yes	362.3	60.25	4.59	0.002
	No	314.9	70.60		
Interest in Academics	Least	336.9	67.11	18.93	0.000
	High	416.26	60.35		

From the above table, there is a significance difference in the score of the male and female students at primary level. The value of t-statistic found as 14.365 with the p-value as 0.000. As the p-value is less than 5%, therefore it is concluded that there is a significant difference between male and female students' performance at primary school level. Also, the value of t-statistic for the comparing the score between the students who were supported and not supported from their parents during the education. The value of the t-statistic found as 0.64 with the p-value as 0.5278. As the significance value is more than 5%, therefore it is concluded that there is no significance difference in the performance of the student having parent's support with those students having no parents' support.

The value of t-statistic for the comparing the score between the students who had tuition facility with those students who do not avail the tuition facility. The value of the t-statistic found as 4.59 with the p-value as 0.002. As the significance value is less than 5%, therefore it is concluded that there is a significance difference in the performance of the student having tuition facility comparing with those having no tuition facility. Similarly, the value of the t-statistic for the students having least interest in education in comparing with those having high interest found as 18.93 with the p-value as 0.000. As P-value is less than 5%, therefore it is concluded that there is a significant difference in the score of the students having high interest in education from those having least interest.

Comparison of Student's Socio-Economic Characteristics

The following table 6 presents the comparison of score in class of student with various characteristics of the students at primary school. It includes comparison of student's score in the class by the various socio-economic characteristics like family type, parent's Education, family size and family income. The table defines the mean and standard deviation of the performance (i.e. score of the student in class). The table compares the score of the student in class with family type and parent's education student t-test. Whereas the score of the student compare among categories of family size and family income using F-statistics. The table also include the p-value for statistical significance.

Table 6: "Comparison of students Performance using t-test**"**

Grouping Variable	Categories	Performance of student in class		t-Statistic	P-Value
		Mean	Standard Dev		
Family Type	Nuclear	398.5	63.26	5.65	0.000
	Joint	382.6	58.89		
Parent's Education	Literate	370.9	67.17	4.064	0.000
	Illiterate	350.2	83.51		

From above table, the value of t-statistic for the comparing the score between their family status. The family status is either is nuclear or joint in nature. The value of the t-statistic found as 5.65 with the p-value as 0.000. As the P-value is less than 5%, therefore it is concluded that there is a significance difference in the performance of the nuclear family student from the joint family student. Similarly, the value of the t-statistic, for the student's preference having educated parents with those of having illiterate parents, found as 4.064 with the p-value as 0.000. As P-value is less than 5%, therefore it is concluded that there is a significant difference in the score of the students having educated parents as compare to those having illiterate parents.

Table 7: "Comparison of Students Performance Using F-statistics**"**

Grouping Variable	Categories	Performance of student in class		F-Statistic	P-Value
		Mean	Standard Dev		
Family Size	Less than 5	359.6	52.36	1.485	0.526
	5-8	340.5	59.68		
	More than 8	336.8	62.30		
Family Income	≤ 40,000	383.38	57.79	61.57	0.000
	40000 - 60000	400.8	62.51		

	≥ 60000	467.07	44.25		

From above table, the value of F-statistic computed for the comparing the score between their family size. The value of F-statistic found as 1.485 with the p-value as 0.526. As the P-value is more than 5%, therefore it is concluded that there is no significance difference in the performance of the various family size. Similarly, table revealed that the value of F-statistic computed for the comparing the score among various family income group. The value of F-statistic found as 61.57 with the p-value as 0.000. As the P-value is less than 5%, therefore it is concluded that there is a significance difference in the performance of students belonging to various income groups at primary schools in district Peshawar.

Comparison of School Characteristics

The following table 8 presents the comparison of score in class of student with various characteristics of school at primary school. It includes comparison of student's score in the class by the various school characteristics like school type, class size and school locality. The table defines the mean and standard deviation of the performance (i.e., score of the student in class). The table compares the score of the student in class with school type, class size and school locality using student t-test. Whereas the score of the student compares among categories of school standard using F-statistics. The table also includes the p-value for statistical significance.

Table 8: "Comparison of Students Performance Using T-test"

Grouping Variable	Categories	Performance of student in class		t-Statistic	P-Value
		Mean	Standard Dev		
School Type	Government	410.19	61.35	10.26	0.000
	Private	376.16	56.97		
Class Size	Less than 40	398.8	67.17	6.64	0.000
	≥ 40	395.7	63.51		
School Locality	Urban	386.3	60.25	6.59	0.000
	Rural	376.9	70.60		

From above table, the value of t-statistic for the comparing the score between their school type, class size and school locality. The school type is either is government and private in nature. The value of the t-statistic found as 10.26 with the p-value as 0.000. As the P-value is less than 5%, therefore it is concluded that there is a significance difference in the performance of government and private primary school student. Also, the score of the students compared between the class size. The value of the t-statistic for the comparing the difference of score found as 6.64 with the p-value as 0.0000. As the p-value is less than 5%, therefore it is concluded that there is significant difference in the score of students between the class size of 40 and more than 40 students. Similarly, the value of t-statistic

for the comparing the score between their school locality. The school locality is either urban or rural in nature. The value of the t-statistic found as 6.59 with the p-value as 0.000. As the P-value is less than 5%, therefore it is concluded that there is a significance difference in the performance of urban and rural primary school locality.

Table 9: "Comparison of Students Performance Using F-test**"**

Grouping Variable	Categories	Performance of student in class		F-Statistic	P-Value
		Mean	Standard Dev		
School Standard	Low	389.6	62.6	87.45	0.000
	Normal	378.5	49.8		
	High	359.8	60.3		

The above table 9 compares the performance of student across the standard of school. The standard of the school categorized as low, normal and high based on facilities. The value of F-statistic computed for the comparing the score among school standard. The value of F-statistic found as 87.45 with the p-value as 0.000. As the P-value is less than 5%, therefore it is concluded that there is a significance difference in the performance of the various standard at primary level based on facilities.

Normality of Response Variable

In this study, the performance of the primary school students being assessed by the marks taken in the class and it is considered as the dependent or response variable. The Normality of response variable assessed by Kolmogorov, Smirnov test. The following table indicates the output of the normality of the marks of the student at primary level.

Table 10: "Normality Test for Score of Students**"**

Null Hypothesis	KS-statistic	P-Value	Decision
Score is Normally distributed	0.831	0.127	Do not reject the Ho:

It is clear from the above table that value of Kolmogorov Smirnov test statistic value found as 0.831 with the p-value as 0.127. As the p-value is more than 5%, therefore the null hypothesis is not rejected and concluded that the score of students of primary school is normally distributed.

Checking the Assumption of Multilevel Modeling

Before fitting the multilevel modeling for assessing the performance of the public and private primary school students, following three main assumptions of multilevel modeling were tested. These assumptions include normality of level one and level two errors, intra cluster correlation and collinearity among explanatory variables. The following subset ion defines the testification of the mentioned assumption one by one.

Normality of level one and level two Errors

The following table 11 defines the results of normality assumption of level one and level two errors. The Kolmogorov Smirnov test was used to test the normality assumption of level one errors because of the large sample while level two errors normality was tested by Shapiro-Wilk test of goodness of fit for distribution. The table defines that value of the test statistic for the both tests along with the P-values. The P-value for Kolmogorov Smirnov and Shapiro-Wilk test obtained as more than 0.05 leading toward the nonrejection of null hypothesis of following the normality assumption. Therefore, it is concluded that both level one and level two errors are approximately normal at 5% level of significance.

Table 11: "Normality Test for Errors at level One &Two"

Errors	Kolmogorov Smirnov		Shapiro-Wilk	
	Test statistic	P-value	Test statistic	P-value
Level One	0.678	0.543	1.025	0.335
Level Two	0.536	0.785	0.875	0.657

Intra Cluster Correlation

Intra Cluster Correlation (ICC) is used to test the homogeneity of individual within cluster. It is the ratio of variation explained by cluster to total variation ranging from 0 to 1. The following table 12 defines ICC of the students socio-economic, students and school characteristics. The results indicate that relation exists between the individuals laying in the same cluster and no close relation found as the value of ICC found less than 50%. The respective value of ICC in the table 4.F ranging from .010 to 0.39, which clearly indicate the violation of independency of observations in classical linear regression model and leads toward the application of the multilevel modeling to the data.

Table 12: "ICC for the various characteristics of Students and school**"**

Predictors	*ICC*
Age	0.368
Gender	0.261
Tuition Facility	0.287
Parents Support	0.321
Interest in Study	0.199
Family Size	0.204
Family type	0.226
Parent Education	0.298
Family income	0.334
School Size	0.301
Class Size	0.367
School Type	0.347
School locality	0.268
School Standard	0.390

Multicollinearity between various characteristics

To check the problem of multicollinearity i.e. the relationship among the explanatory variables, correlation coefficient determined between the students, schools and socio-economic characteristics regarding students of Primary school at district Peshawar. The following tables indicates the relationship among explanatory variables regarding various characteristics of school and student respectively.

The following table 13 represents the correlation among the different characteristics at student level. The table defines that there exists almost very low relationship or negative relationship among the variables. The table explains that age has a negative relationship with gender and tuition facility taken by the student while a positive very rare relationship with parents support and interest in study. Also, gender of the student showed a rare positive relationship with parents support and tuition facility while a negative relationship with interest in education of student. Similarly, parents support showed a positive relationship with tuition facility and interest in education while the tuition facility availed by the student showed a rare relationship with interest in education of primary school student at district Peshawar.

Table 13: "Correlation Matrix of Students' Characteristics**"**

Variables	Age	Gender	Parent Support	Tuition Facility	Interest in education
Age	1	- 0.0541	0.0158	-0.0137	0.0540
Gender		1	0.081	0.0233	-0.089
Parent support			1	0.0591	.0478
Tuition Facility				1	0.0091
Interest in Education					1

The following table 14 represents the correlation among the different socio-economic characteristics at student level. The table defines that there exists almost very low relationship or negative relationship among the variables. The table explains that family type has a positive relationship with family size, parent's education and family income. The table also shows that family size has negative relationship with parent's education while a positive relationship with income of family. Also, the parent's education showed ap positive relationship with family income of students at primary school level in district Peshawar.

Table 14: "Correlation Matrix of Socio-economic Characteristics**"**

Variables	Family Type	Family Size	Parent Education	Family Income
Family Type	1	- 0.0428	0.0400	0.1926
Family Size		1	0.2140	0.3273
Parent Education			1	0.0457
Family Income				1

The following table 15 represents the correlation among the different school characteristics at primary level. The table defines that there exists almost very low relationship or negative relationship among the variables. The table explains that school type has a positive relationship with school size, class size and school standard and negative relationship with school locality. The table also shows that school size has positive relationship with class size and school locality while a negative relationship with school standard. Also, the class size showed a positive relationship with school locality and school standard of students at primary school level in district Peshawar.

Table 15: "Correlation Matrix of School Characteristics**"**

Variables	School Type	School Size	Class Size	School Locality	School Standard
School Type	1	0.2510	0.1218	-0.1887	0.2280
School Size		1	0.3860	0.0897	-0.0962
Class Size			1	0.1823	0.2062
School Locality				1	0.0610
School Standard					1

Best Model Selection

To select the best fit model for explaining the performance of the primary school student at public and private sector level, three different approaches were used to select the best model. These approaches include Deviance, Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). The model being split into level one and level 2 based on set of explanatory variables. Forward selection method was used to select the best model explaining the performance of the students.

Public/Government Sector Schools

The table 16 in Appendix-I, illustrate the three different model selection criteria for the public sector primary schools at district Peshawar. The forward section method being applied to select best model explaining the performance of the primary school students. Each model obtained by forward selection method, being evaluated by Deviance (D), AIC and BIC. The table indicates that base model has D = 6852.01, AIC = 6588.03 and BIC = 6406.98. The best model criteria require minimum D, AIC and BIC. When the age as independent variable added to model, the value of D, AIC and BIC decreases indicating that age as important explanatory variable and must be included in model. The same position observed when students' characteristics like gender, interest in study, tuition facility availed, and locality of student added to model.

At level II the school characteristics added, and effect noted. As the school size added to model, the resultant D, AIC and BIC observed as 6581.8, 6589.6 and 6612.3 respectively, which observed more than that of base model. Different multilevel model considered at student characteristics i.e. level-I and school characteristics i.e. level-II and evaluated at mentioned criteria. It includes two factors combination, three factor models, four factors model, five factor model and six factor combination model. The final model includes all factors related to student's characteristics (at level-I) and two factors related to school characteristics at level-II. In the table the model containing best representative combination of the factors are being highlighted and attached as appendix I.

Best Selected Model

The best selected model contains all six variables as an explanatory which explains the score or performance of the primary school student at district Peshawar. These includes age, gender, interest in study, parent support, tuition facility availed, living area of student at level I whereas level-II includes two variables i.e. class size and school standard. The following table 17 illustrate the result of Pseudo R^2 for final fitted model containing fixed and random component of the regression coefficient at level I and level II

Table 17: "Pseudo R-square for Final Model**"**

Pseudo R^2	%
Level-I	74.025
Level-II	64.34

The values of the Pseudo R^2 indicates the goodness of fit of regression model. The value of level-I indicates that 74.02 % variations in the performance of the student at primary level schools are being explains by the variables at lower level included in model while 64.34 % variations being explains by variables included in upper level if model.

The Estimated Regression Model for Public Sector Schools

The estimated multilevel regression model for the public or government primary schools is

$$\begin{aligned} Score_{Public} = {} & 265.857 - 0.1257\,(Age) - 0.548\,(gender_{male}) \\ & + 0.4868\big(Interested_{yes}\big) + 1.056\,(Locality_{Urban}) \\ & - 0.625\,\big(Tuition_{yes}\big) + 0.252\,\big(Parent\ Support_{yes}\big) \\ & - 0.235\,(Class\ Size) + 0.810\,(School\ Standard_{Normal}) \\ & + 1.745\,(School\ Standard_{Good}) \\ & + 0.0008\,(School\ Std * Interest\ in\ Study) \\ & - 0.245\,(Class\ Size * Interest\ in\ Study) \end{aligned}$$

In above model, the average marks of the students in the public sector primary schools will is 265.87 when all the factors have zero effect. The model explains that for the change in gender has a negative impact on the performance of the students at primary level and will decrease by 0.1257 units. The model indicates that interest in education has shown a positive impact on score of the public sector primary schools. It defines that foe the unit increase in the interest of the student; the score will increase by 0.4868 units. Similarly, locality of the student also has a positive relationship with the score of the public sector

primary schools. It is found that for a unit increase in locality, the sore of the student will increase by 1.056 units. The tuition facility availed by the public sector primary school students showed an inverse relationship with the score of the students. it is observed that if the tuition facility changes by one unit, the score of the student will decrease by 0.625 units.

It is also found that parents support their child showed a positive impact on the performance of the student in primary level. It is found that increase in parents support will also increase the score of the student by 0.252 time. At level-II, the factor class size showed a negative impact on the performance of the student. It is found that for a unit increase in the class size, the performance of the student will decrease by 0.235 times. School standard showed a positive relationship with the score of the public sector primary schools. It is found that if normal standard government primary schools improve their standard, the sore of the student will increase by 0.810 units. Similarly, School good standard showed a positive relationship with the score of the public sector primary schools. It is found that if good standard government primary schools improve their standard, the sore of the student will increase by 0.810 units

The model also contains the effect of interaction terms on the score of the government sector primary schools. The model defines that joint or interaction effect of school standard and student interest in the education have a positive impact on the score or performance of the student. It is observed that for at increase in joint effect of school standard and student interest in education, the performance of the student will increase by 0.0008 units. Also, the class size and the school's standard have showed a negative impact on the performance of the student. It indicates that if the standard schools have overloaded class size it will decrease the performance of the student in public sector primary schools.

Private Schools

The table 18 in Appendix-II illustrate the three different model selection criteria for selecting the model which explains the performance of the primary school in private sector. This model contains nine variables i.e. four at level I and four at level II. Level-I four variables are gender, family type, family monthly income, tuition facility availed and interest in study. Whereas the variables at level-II are school standard, school Size, Class size and location of school. The forward section method being applied to select best model explaining the performance of the primary school students. Each model obtained by forward selection method, being evaluated by Deviance (D), AIC and BIC. The table indicates that base model has D = 6852.01, AIC = 6588.03 and BIC = 6406.98. The best model criteria require minimum D, AIC and BIC. When the gender as independent variable added to model, the value of D, AIC and BIC decreases indicating that age as important explanatory variable and must be included in model. The same position observed when students' characteristics like interest in study, family type, tuition facility availed, and locality of student added to model.

At level II the school characteristics added, and effect noted. As the school size added to model, the resultant D, AIC and BIC observed as 6581.8, 6589.6 and 6612.3 respectively, which observed more than that of base model. Different multilevel model

considered at student characteristics i.e. level-I and school characteristics i.e. level-II and evaluated at mentioned criteria. It includes two factors combination, three factor models, four factors model, five factor model and six factor combination model. The final model includes all factors related to student's characteristics (at level-I) and two factors related to school characteristics at level-II. In the table the model containing best representative combination of the factors are being highlighted and attached as appendix II.

CONCLUSION AND RECOMMENDATIONS

To determine the factors regarding the students' performance at the primary level in schools in district Peshawar, primary data was collected from students of the school and the schools of the students, respectively. The data was collected from the students of various gender of the students at level-I. At level I, there were 1275 male students and 1289 female students with the percentage of 49.7 and 50.3 respectively from the primary schools of district Peshawar. At level-II, there were 50 male and 50 female students from the Primary schools. The student of primary schools belonged to nuclear and joint family setup. The student's score in the class by the various characteristics like gender of students, parent support, availing tuition facility and students in the academic. The study showed a significance difference in the score gender wise, no significance difference by parental support, significant difference in the score interest in education. Also, a significance difference in the performance by having tuition facility a significance difference in score by family structure and a significant difference in the score observed by parent's education. (Seltzer, M. H. 1994: Lüdtke et al 2009).

In this study, the performance of the primary school students being assessed by the marks taken in the class and it is considered as the dependent or response variable and found normal in nature, which was assessed by Kolmogorov, Smirnov test. The multilevel regression model was fitted for government and private school students score respectively. The study showed that score of the student at Public School primary level in district Peshawar has positive with interest in education, locality of student, parents support, school standard and school standard with interest while a negative relation with age, gender, tuition facility and class size with student interest in education. The study showed that score of the student at private school primary level in district Peshawar has positive with interest in education, family type, monthly family income, school standard and school standard with interest while a negative relation with gender and class size.

REFERENCES

Boyd, L.H., Iversen, G.R. (1979). Contextual analysis: Concepts and statistical techniques. Wadsworth Publishing Company.

Berkhof, J., Snijders, T.A. (2001). Variance component testing in multilevel models. *Journal of Educational and Behavioral Statistics*, 26(2), 133-152.

Cohen, M.P. (1998). Determining sample sizes for surveys with data analyzed by hierarchical linear models. *Journal of Official Statistics*, *14*(3), 267.

Draper, N.R., Smith, H. (1998). *Applied regression analysis (Vol. 326)*. John Wiley & Sons.

DiPrete, T.A., Grusky, D.B. (1990). The multilevel analysis of trends with repeated cross-sectional data. *Sociological Methodology*, 337-368.

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1-22.

De Leeuw, J., Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, 11(1), 57-85.

Darling, N., Caldwell, L.L., Smith, R. (2005). Participation in school-based extracurricular activities and adolescent adjustment. *Journal of leisure research*, *37*(1), 51-76.

Goldstein, H. (1986). Efficient statistical modelling of longitudinal data. *Annals of human biology*, *13*(2), 129-141.

Goldstein, H., Spiegelhalter, D.J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *159*(3), 385-409.

Fielding, A., Yang, M., Goldstein, H. (2003). Multilevel ordinal models for examination grades. *Statistical modelling*, *3*(2), 127-153.

Hüttner, H.J., van den Eeden, P. (1995). The multilevel design: A guide with an annotated bibliography, 1980-1993 (No. 23). ABC-CLIO.

Hartley, H.O., Rao, J.N. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, *54*(1-2), 93-108.

Kreft, I.G., Kreft, I., de Leeuw, J. (1998). Introducing multilevel modeling. Sage.

Krejcie, R.V., Morgan, D.W. (1970). Determining sample size for research activities. *Educational and psychological measurement*, *30*(3), 607-610.

Lindley, D.V., Smith, A.F. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(1), 1-18.

Raudenbush, S., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of education*, 1-17.

Televantou, I. (2015). Phantom effects in school composition research: Consequences of failure to control biases due to measurement error in traditional multilevel models. *School Effectiveness and School Improvement*, *26*(1), 75-101.

Toby, J. (1957). The differential impact of family disorganization. *American sociological review*, *22*(5), 505-512.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (Vol. 1). sage.

Seltzer, M. H. (1994). Studying variation in program success: A multilevel modeling approach. *Evaluation Review, 18*(3), 342-361.

Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary educational psychology, 34*(2), 120-131.

Luke, D. A. (2019). Multilevel modeling (Vol. 143). Sage publications.

Lazega, E., & Snijders, T. A. (Eds.). (2015). Multilevel network analysis for the social sciences: Theory, methods and applications (VOL: 12). Springer.

Tate, R.L., Hokanson, J.E. (1993). Analyzing individual status and change with hierarchical linear models: Illustration with depression in college students. *Journal of Personality, 61*(2), 181-206.

Appendix-I

Table 16: "Best model Selection Criteria for Government Sector Primary School"

Variables	Level	Model	D	AIC	BIC
Base model	1 & 2	No Variables	6852.01	6588.03	6406.98
One	1	Age (A)	6488.6	6496.6	6519.3
	1	Gender (G)	6526.5	6534.65	6557.29
	1	**Interest (IS)**	**6327.15**	**6339.15**	**6373.11**
	1	P Support (PS)	6579.99	6587.99	6610.63
	1	Tuition (T)	6487.65	6495.65	6518.29
	1	Locality (L)	6378.66	6386.66	6409.34
	2	School type (ST)	6579.38	6587.38	6609.99
	2	School Size (SZ)	6581.66	6589.66	6612.3
	2	Class size (CS)	6567.8	6575.8	6598.44
	2	**School Std (SS)**	**6509.77**	**6519.77**	**6548.01**
Two	1	A+G	6426.4	6436.4	6464.7
	1	**A+IS**	**6223.2**	**6235.2**	**6269.2**
	1	A+L	6298.45	6308.45	6336.75
	1	G+IS	6282.68	6294.68	6328.64
	1	G+PS	6225.39	6535.39	6563.69
	1	G+L	6328.79	6338.79	6367.1
	2	ST+SS	6578.1	6588.1	6616.38
	2	ST+CS	6576.43	6586.43	6614.7
	2	ST+SS	6508.83	6520.83	6554.79
	2	**CS+SS**	**6508.3**	**6520.3**	**6554.3**
Three	**1**	**A+G+IS**	**6173.2**	**6189.2**	**6234.4**
	1	A+G+PS	6425.2	6437.2	6471.1
	1	A+G+L	6242.6	6254.6	628.62
	1	G+IS+PS	6282.8	6298.8	6343.7
	1	G+IS+L	6088.8	6104.8	6150.1

	2	ST+SS+CS	6572.1	6584.1	6617.9
	2	ST+SS+SD	6507.4	6521.4	6560.8
	2	SS+CS+SD	6508.1	6522.1	6561.7
Four	1	A+G++IS+PS	6172.8	6190.8	6241.1
	1	**A+G+IS+L**	**5993.5**	**6011.55**	**6062.5**
	1	G+IS+PS+T	6148.7	6166.7	6217.6
	1	G+IS+PS+L	6088.3	6106.4	6157.3
	1	A+G+PS+T	6357.3	6371.3	6410.4
	2	ST+SS+CS+SD	6507.1	6523.1	6568.3
Five	1	A+G+IS+PS+T	6068.2	6088.2	6144.7
	1	**A+G+IS+PS+L**	**5993.3**	**5913.3**	**6069.7**
Six	**1**	**A+G+IS+PS+T+L**	**5884.50**	**5906.5**	**5968.8**

Appendix-II

Table 18: "Best model selection Criteria for Private Primary Schools**"**

Variables	Level	Model	D	AIC	BIC
Base model	1 & 2	No Variables	6852.01	6588.03	6406.98
One	1	Family Type	6545.2	6553.3	6575.6
	1	Gender (G)	6526.5	6534.65	6557.29
	1	**Interest (IS)**	**6327.15**	**6339.15**	**6373.11**
	1	P Support (PS)	6579.99	6587.99	6610.63
	1	Tuition (T)	6487.65	6495.65	6518.29
	2	School type (ST)	6579.38	6587.38	6609.99
	2	School Size (SZ)	6581.66	6589.66	6612.3
	2	Class size (CS)	6567.8	6575.8	6598.44
	2	**School Std (SS)**	**6509.77**	**6519.77**	**6548.01**
Two	1	FT+G	6426.4	6436.4	6464.7
	1	**IS + FMI**	**6223.2**	**6235.2**	**6269.2**
	1	G+IS	6282.68	6294.68	6328.64
	1	G+PS	6225.39	6535.39	6563.69
	1	G+L	6328.79	6338.79	6367.1
	2	ST+SS	6578.1	6588.1	6616.38
	2	ST+CS	6576.43	6586.43	6614.7
	2	ST+SS	6508.83	6520.83	6554.79
	2	**CS+SS**	**6508.3**	**6520.3**	**6554.3**
Three	1	FT+G+FMI	6173.2	6189.2	6234.4
	1	G+FT+IS	6425.2	6437.2	6471.1
	1	**FT+FMI+IS**	**6507.4**	**6521.4**	**6560.8**
	2	ST+SS+CS	6508.1	6522.1	6561.7
Four	**1**	**G+FT+FMI+IS**	**6176.6**	**6198.68**	**6260.3**