Final Project Report

OMIS-645

New York State High School Quality Statistical Analysis (2014-2015)

By

Darshan Kumar Saravanan (Z1947418)

Ritesh Boda (Z1857245)

Durgaprasad Koppala (Z1981777)

Sujauddin Syed (Z1969480)

Under the Guidance of Prof. Kishen Iyengar

OM&IS, Northern Illinois University

Introduction

The School Quality Dataset shows the data of more than 450 high schools in New York and various factors such as, Effect of Leadership, %Trust, % Supportive Environment, Economic Need Index, Years of experience of teachers, Years of Principal at the same school, Student Achievement Rate, %Rigorous Instruction etc.

This Dataset is found to be reliable as the source is from a government website of USA. So, we further cleaned the dataset by replacing N/A values blank for our analysis using SAS.

We analyzed the dataset to gain a deeper understanding of the factors that effects Leadership, Quality of Curriculum, Student Achievement Rate etc. By using different techniques of Statistical Analysis using SAS tool.

Analysis

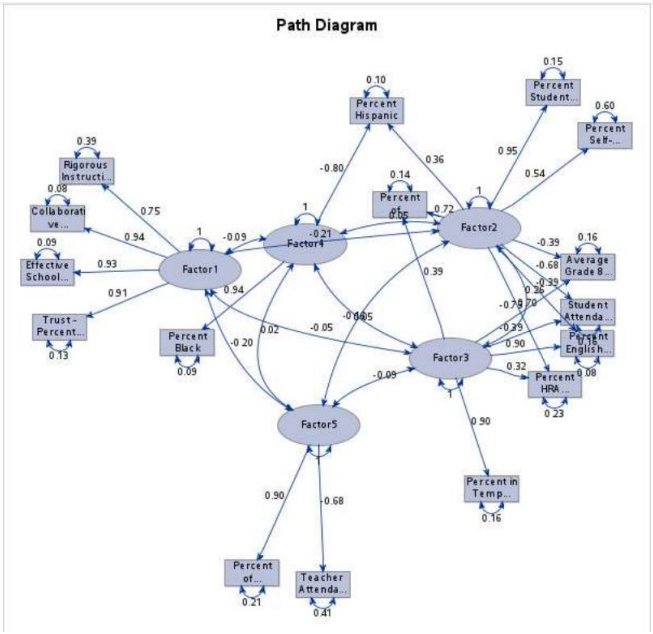
Note: We conducted all the analysis with a 95% confidence level.

1) **What factors influence chronic absenteeism in the schools of New York State? How can we limit absenteeism and increase student attendance in schools?**

   Independent Variables Included:

   To look at the variables that influence chronic Absenteeism in the schools of New York. We used variables such as:

   -Average Grade 8 Proficiency.
   -Economic Need Index
   -Years of Principal experience at this school
   -Teacher's attendance rate.
   -Moreover, we used a factor (School & Instructional Quality) in this Analysis.



-The Factor included in this Analysis consists of the Variables as follows:

1. Collaborative Teachers Percentage
2. Effective School Leadership Percentage
3. Trust Percentage
4. Rigorous Instruction Percentage

- First, we did a factor analysis on all the numerical variables present in the dataset to reduce it to Factors.
- Then we found significance for Factor1 alone which I would name as School and Instructional Quality.

Model: MODEL1
Dependent Variable: Percent of Students Chronically Percent of Students Chronically Absent

Number of Observations Read	500
Number of Observations Used	441
Number of Observations with Missing Values	59

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	7.65047	1.53009	142.10	<.0001
Error	435	4.68383	0.01077		
Corrected Total	440	12.33431			

Root MSE	0.10377	R-Square	0.6203
Dependent Mean	0.37084	Adj R-Sq	0.6159
Coeff Var	27.98148		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Variance Inflation
Intercept	Intercept	1	1.93394	0.48326	4.00	<.0001	0	0
Average Grade 8 Math Proficiency	Average Grade 8 Math Proficiency	1	-0.19116	0.01791	-10.67	<.0001	-0.45025	2.03841
Economic Need Index	Economic Need Index	1	0.36833	0.04517	8.15	<.0001	0.34082	2.00102
Years of principal experience at	Years of principal experience at this school	1	-0.00358	0.00136	-2.62	0.0090	-0.07807	1.01431
Factor1		1	-0.01903	0.00529	-3.60	0.0004	-0.11374	1.14532
Teacher Attendance Rate	Teacher Attendance Rate	1	-1.38964	0.49979	-2.78	0.0057	-0.08759	1.13670

After running the Analysis:

-The overall Model is Significant
-Adjusted R Square – 61.59%, so the model is a good fit.
-All our Independent Variables are significant as well.

Our Biggest predictor of the Model is Average grade 8 Proficiency.

- For Every unit increase in the Average grade 8 Proficiency, chronic absenteeism decreases. Average grade 8 Proficiency has a negative effect on Chronic absenteeism.

  -meaning as the average math proficiency of a school increases, the chronic absenteeism decreases.

- For Every unit increase in the Economic need index, chronic absenteeism Increases.
  - Meaning, as the economic need for a school increase, chronic absenteeism increases.
- For every unit increase in the years of experience for a principal at a school, chronic absenteeism decreases.
  - Meaning, the chronic absenteeism in a school decrease, when the principal stays longer in a school.
- For Every unit increase in the Teacher Attendance Rate, chronic absenteeism decreases.
  - Meaning, chronic absenteeism decreases when the teacher's attendance rate is higher.
- For every unit increase in (School & Instructional Quality) the Factor 1(i.e., combination of % of collaborative teachers, % of effective school leadership, % trust, and % Rigorous Instruction), chronic Absenteeism decreases.

Chronic Absenteeism = 1.93394 -0.19116 (Average grade 8 Math Proficiency) + 0.36833 (Economic needs Index) – 0.00358(Years of Principal experience at this school) – 0.01903(Factor 1) -1.38964 (Teacher Attendance Rate)

Assumptions:

- The Residuals are normally Distributed.
- The Residuals are homoscedastic.
- There is no autocorrelation, as it is not a time series Data.

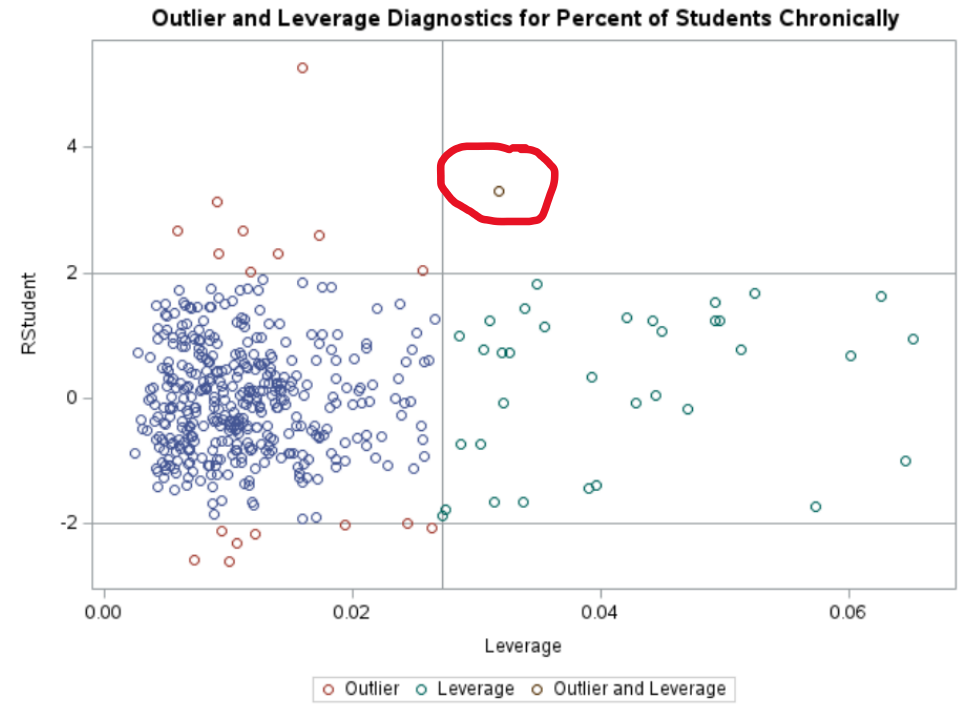
Outlies and Leverage Points:

- Observation 124 must be removed, as it is an outlier and leverage point, and is affecting our analysis.

-Observation 124 is Samuel Gompers Career and Technical Education High, Bronx, NY.

-This is an Outlier and a leverage point as the enrollment number in is school is only 78.

- This is an outlier as the observed value is completely different from the predicted value and this might be because the variation in the predicted value is being affected by all the variables used in the analysis.



- So, from this Analysis, I conclude that Schools must focus more on Math classes to make it Interesting, Satisfy the economic needs, maintain a principal that stays in the school

longer, and have a good teacher attendance rate to dampen the Chronic absenteeism in the schools of Newyork State.

So, how can we limit absenteeism, while increasing student attendance in schools?

Using the results from the above analysis, further we explored a little deeper to identify the variables that help In Increasing the Student's Attendance Rate.

- While analyzing the variables that influence absenteeism in the schools of Newyork.
- We found out that chronic absenteeism strongly correlates with the students' Attendance Rate.
- We thought of moderating this relationship between chronic absenteeism and Attendance rate using a moderator that would Increase the Students Attendance Rate.

- Interestingly, Average Grade 8 Proficiency moderates the relationship between Chronic Absenteeism and Students Attendance Rate.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1.70153	0.56718	2315.02	<.0001
Error	486	0.11907	0.00024500		
Corrected Total	489	1.82060			

Root MSE	0.01565
Dependent Mean	0.87589
R-Square	0.9346
Adj R-Sq	0.9342
AIC	-3578.00407
AICC	-3577.88010
SBC	-4053.22645

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.024047	0.008185	125.11	<.0001
Percent of Students	1	-0.555224	0.031969	-17.37	<.0001
Average Grade 8 Math	1	-0.013687	0.002946	-4.65	<.0001
Percent o*Average Gr	1	0.102421	0.015337	6.68	<.0001

2) Does student achievement depend on the quality of how interesting and challenging the curriculum of a school in New York state is?

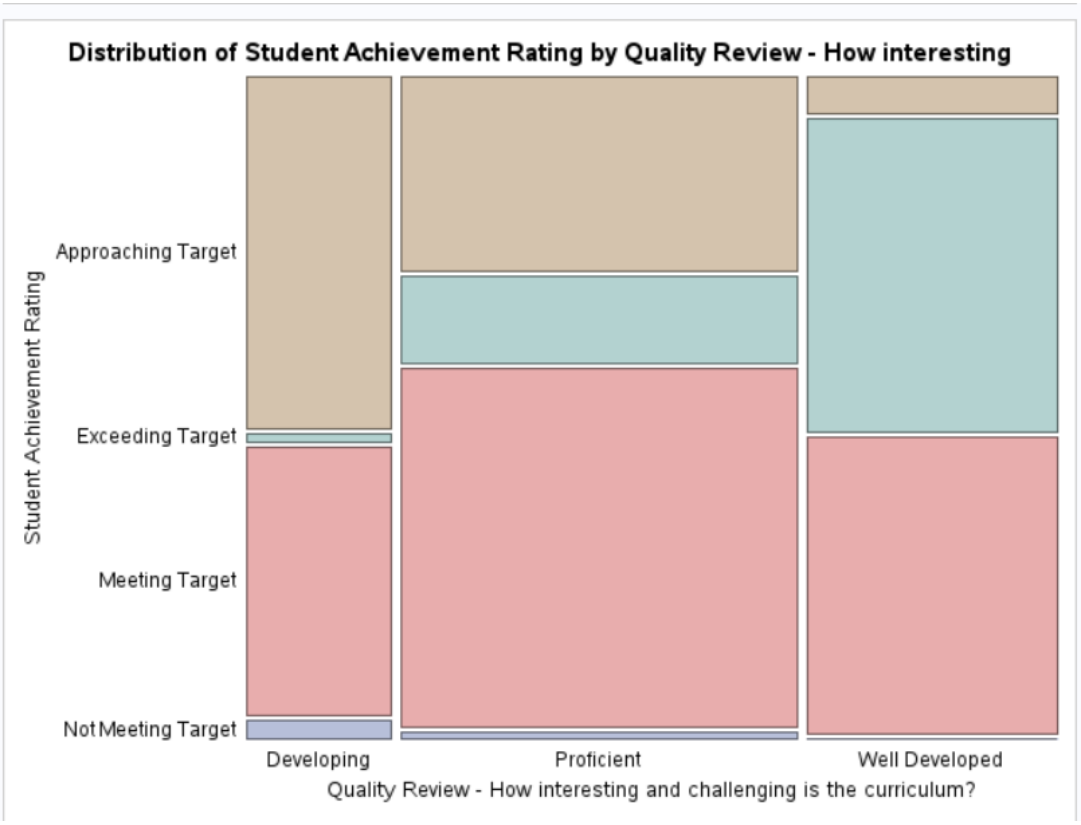
Null Hypothesis – Student achievement does not depend on the quality of how interesting and challenging the curriculum of a school in New York state is.

Alternative Hypothesis - Student achievement depends on the quality of how interesting and challenging the curriculum of a school in New York state is.

- By looking at the P-value of the Chi-square value, we conclude that both the variables are dependent.

- By looking at the Distribution of Student Achievement Rating by Quality review, we can say that:

1) In Schools that have Well developed curriculum, most of the student's achievement ratings are exceeding the Target.
2) In Schools that have Proficient curriculum most of the student's achievement ratings are Meeting the Target.
3) In Schools that have Developing curriculum most of the student's achievement ratings are Approaching the Target.



3) Factors Explaining the Trust of Schools in New York State

Introduction:

The purpose of this study is to determine the factors that contribute to the variation in trust among schools in New York State. The study uses both numerical and categorical variables to analyze the trust of schools. The aim of this report is to provide a summary of the findings from the study.

Methodology:

The study utilized numerical and categorical variables to analyze the trust of schools. The categorical variables include ratings with categories such as Not Meeting Target, Meeting Target, Approaching Target, and Exceeding Target. The study analyzed the significant variables contributing to the variation in trust among schools. The variables were:

- Rigorous Instruction
- Collaborative Teaching
- Supportive Environment
- Economic Need Index
- Percent of Teachers with 3+ years of experience
- Strong Family-Community ties (Categorical)

Model: MODEL1
Dependent Variable: Trust - Percent Positive Trust - Percent Positive

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Variance Inflation
Intercept	Intercept	B	0.26696	0.02776	9.62	<.0001	0	0
Rigorous Instruction	Rigorous Instruction	1	-0.12564	0.03220	-3.90	0.0001	-0.12173	1.95401
Collaborative Teache	Collaborative Teache	1	0.52650	0.02239	23.51	<.0001	0.74462	2.01327
Supportive Environme	Supportive Environme	1	0.26779	0.02783	9.62	<.0001	0.26950	1.57497
Economic Need Index	Economic Need Index	1	0.01835	0.00817	2.25	0.0252	0.05576	1.23647
Percent of teachers	Percent of teachers	1	0.01691	0.00770	2.20	0.0287	0.05232	1.13926
Strong Family-Commun Approaching	Strong Family-Commun Approaching Target	B	0.02299	0.00445	5.17	<.0001	0.21820	3.57885
Strong Family-Commun Exceeding T	Strong Family-Commun Exceeding Target	B	0.03528	0.00602	5.86	<.0001	0.19601	2.24490
Strong Family-Commun Meeting Tar	Strong Family-Commun Meeting Target	B	0.02997	0.00468	6.40	<.0001	0.28313	3.92341
Strong Family-Commun Not Meeting	Strong Family-Commun Not Meeting Target	0	0	.	.	.	.	.

Results:

The model used to analyze the data was statistically significant and explains approximately 81.02% of the variance in the dependent variable "Trust - Percent Positive" based on the

independent variables included in the model. The findings show that collaborative teaching is the highest predictor of trust among schools in New York State. For every one-unit increase in Collaborative Teaching, Trust - Percent Positive increases. The other significant variables that contribute to the variation in trust among schools are Supportive Environment, Economic Need Index, and Percent of Teachers with 3+ years of experience.

Interestingly, the findings show that Rigorous Instruction has a negative impact on the trust of schools. For every one-unit increase in Rigorous Instruction, Trust - Percent Positive decreases**.**

Assumptions:

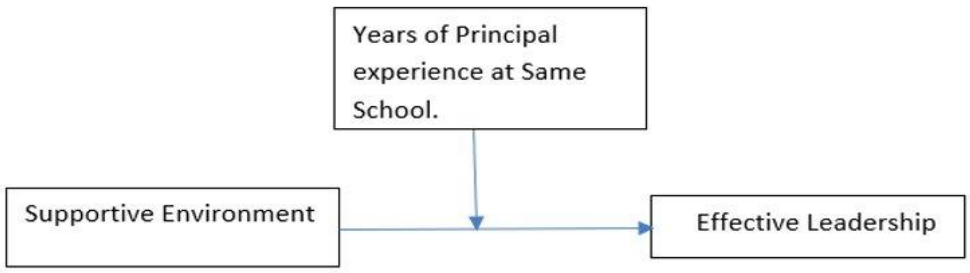
The study is based on three assumptions; the residuals are normally distributed, the residuals are homoscedastic, and there is no autocorrelation since it is not a time series data.

Conclusion:

In conclusion, to improve trust in schools in New York State, the study recommends improving collaborative teaching and the supportive environment of schools. Other factors, such as rigorous instruction, economic need index, and experienced teachers, also have an impact on trust, but to a lesser degree. The findings from the study provide valuable insights for policymakers and educators to enhance the trust of schools in New York State.

4) Does Years of Principal at the same school moderate the relationship between Supportive Environment and Effective School Leadership?

To analyze the above the situation, Firstly I would check the type of impact Supportive Environment is making on Effective School Leadership.



Independent Variable: Supportive Environment (%)

Dependent Variable: Effective School Leadership (%)

We run this analysis using Linear Regression,

Main Model Effect:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.03017	1.03017	131.62	<.0001
Error	489	3.82738	0.00783		
Corrected Total	490	4.85755			

Root MSE	0.08847	R-Square	0.2121
Dependent Mean	0.81945	Adj R-Sq	0.2105
Coeff Var	10.79627		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.09345	0.06341	1.47	0.1412
Supportive Environment - Percent	Supportive Environment - Percent Positive	1	0.87704	0.07645	11.47	<.0001

The above result states that the model is significant, and the variable Supportive environment can predict/effect the School Leadership. Also, it shows that 21% of variance in effective school leadership is explained by Supportive environment in that school.

Cross Product Effect:

Now I would like to see if the variable “number of years Principal at the same school” can affect the relationship between Supportive environment and Effective School Leadership.

To run this analysis of Moderation,

Moderator: Number of Years Principal at same School (No. Of Years)

Independent Variable: Supportive Environment (%)

Dependent Variable: Effective School Leadership (%)

We run a Linear Regression with the cross product of Moderator and Independent Variable.

Least Squares Model (No Selection)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.99396	0.33132	42.26	<.0001
Error	439	3.44203	0.00784		
Corrected Total	442	4.43599			

Root MSE	0.08855
Dependent Mean	0.81883
R-Square	0.2241
Adj R-Sq	0.2188
AIC	-1698.87565
AICC	-1698.73835
SBC	-2127.50137

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.082335	0.108934	-0.76	0.4502
Supportive Environme	1	1.093488	0.132012	8.28	<.0001
Years of principal e	1	0.035174	0.017262	2.04	0.0422
Supportiv*Years of p	1	-0.043313	0.020762	-2.09	0.0375

From the P values generated, we understand the model is significant and we see that there is a negative effect on the dependent variable because of the moderator.

So, the effect of school leadership is being dampened by this moderator, though there is a good percent of supportive environment.

Conclusion: The analysis suggests that there should not be a constant Principal in a school for a longer period of time for effective Leadership.

References

"2014-2015 School Quality Reports Results for High Schools." *Data.gov*, Publisher Data.cityofnewyork.us, 29 Apr. 2023, https://catalog.data.gov/dataset/2014-2015-school-quality-reports-results-for-high-schools.