Analyzing the Impact of School Climate Including Violence, Motivation, Safety, and Overall Atmosphere on Academic Performance in Israel

382.1.2601 | Introduction to Data Science | June 2025

Amit Lalush, Dor Benita, Moran Shavit, Yoav Evron

Link to Our Github Repository | Link to the Data (June 2025)

1. Introduction

Students' academic achievement is influenced by many factors, such as teaching quality, socioeconomic status, and parental involvement. One important factor is the school climate - the way students experience the emotional and social atmosphere at school.
In this project, we ask: which specific aspects of the school climate are most **strongly connected** to students' math grades? And do these connections change between different socioeconomic groups?

This question is important because even though schools understand that the school climate affects student achievement, they don't always know which specific areas to improve. As former students in the Israeli school system, this topic is also meaningful to us personally. It is also a complex question, since many school climate factors are connected to each other, and also to students' socioeconomic background[1].

Previous studies show that when students feel safe, have good relationships with teachers, and feel like they belong, they usually do better in school[2]. On the other hand, when students face bullying, violence, or fear, their grades often go down[3].
Some studies point out that just reducing violence is not always enough to help students succeed.

To explore the relationship between school climate and students' academic performance, we will use descriptive statistics, heatmaps, and regression models to identify the climate variables most strongly associated with math achievement.
We hypothesize that negative emotional climate indicators (such as fear or bullying) will predict lower math scores, while positive indicators (such as belonging and teacher encouragement) will be linked to better performance.
Furthermore, we will test whether these patterns differ between socioeconomic groups. By comparing results for students from different backgrounds, we hope to understand whether school climate can help reduce inequality, as some studies suggest.

Our unique contribution is identifying which specific aspects of school climate are most strongly linked to students' math achievement in the Israeli context. In addition, we compare how these relationships differ across socioeconomic groups, helping to untangle the common overlap between school climate and socioeconomic status.

2. Data Overview

Our project relies on three public datasets retrieved from the RAMA "Meitzav" data portal. These datasets include:

i. Scores by schools – Annual "Meitzav" test results per school.
ii. Climates by school – Responses to teacher and pupils surveys assessing school climate.
iii. Metadata on schools – Metadata about each school's sector, location, and socioeconomic level.
Each dataset is described in detail in the README file.

Each record in the dataset represents one school in one academic year, and includes both academic performance and climate survey responses. In total, the dataset contains information for approximately 5,900 school-year observations, spanning grades 5 and 8 between 2008–2020.
The dataset includes three main types of features.

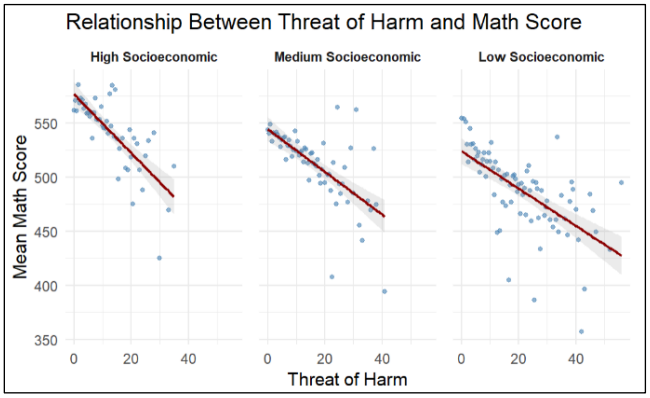
3. Methods and results

Data Preprocessing and Structure

The datasets were merged using the semel_mosad (school identifier) and year columns. Climate survey responses were pivoted to wide format, turning each var_id into a distinct feature. We identified numerous var_id columns with over 57% missing values. Instead of imputation, we excluded these columns. We then kept only columns with over 80% non-null values, reducing missing data significantly (under 0.7%). The remaining few rows with missing values (under 0.01% of total data) were removed using drop_na(), including any unrecorded scores. Finally, we split the cleaned data into three subsets based on socioeconomic status (SES): low (3), mid (2), and high (1), for independent analysis.

Exploratory Data Analysis

We began by computing correlation heatmaps between all school climate variables and math scores. Positive correlations consistently appeared for variables related to satisfaction, belonging, and emotional safety (e.g., "I enjoy going to school", "I feel good at school"), while negative correlations emerged for variables related to fear and violence (e.g., "A pupil threatened to hurt me"). These patterns were stable even when analyzing each SES group separately, indicating that the relationship between school climate and performance persists within different socioeconomic contexts.

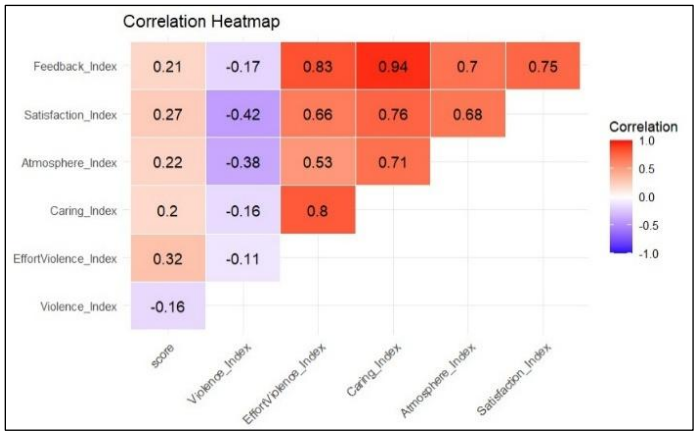


Next, we delved deeper into the data, examining trends across different years. We observed significant changes in both scores and variables within each SES layer over the years. This led us to normalize each column by the annual average socioeconomic score. Additionally, we found very strong correlations among variables within the same categories (e.g., violence, satisfaction). Therefore, we chose to group these variables into distinct categories: feedback, satisfication, atmosphere, caring, effort to prevent violence, and actual violence.

Predictive Model and Feature Importance

For each SES group, we split the data into training (80%) and testing (20%) sets. To quantify the relationship between school climate and mathematics performance, and to identify the most correlated predictors, we trained a separate linear regression model for each SES group using the respective training set.

Given the high correlation within the variables (as shown in the heatmap, we removed the feedback variable because it has too high correlation with another more interesting variable - caring. And to assess the stability of the model coefficients due to the fact the correlation in still pretty high between the variables, we performed a bootstrap analysis with 100 resamples per SES group, allowing us to estimate confidence intervals for each predictor's effect on math scores.

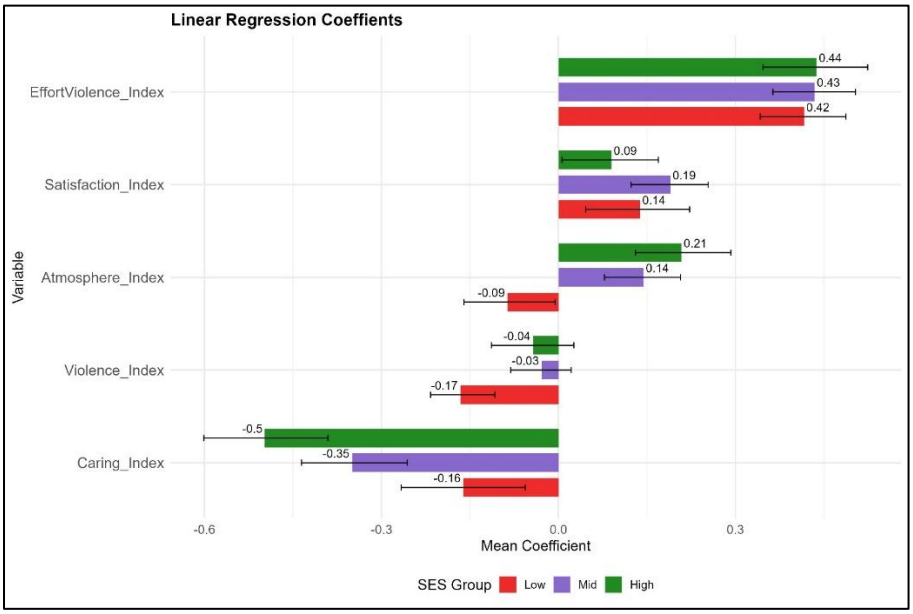


Model evaluation was conducted using 5-fold cross-validation for each SES group, with a new 80/20 train-test split in each fold. Since all variables were standardized, the results are reported in terms of standard deviations (SD) of the distribution of the test data values.

$Socioeconomic\ group$	R^2	$MAE\ (in\ SDs)$	$RMSE\ (in\ SDs)$
$High$	0.10	0.70	0.90
$Medium$	0.16	0.70	0.91
Low	0.18	0.70	0.92

the low r^2 indicates that climate variables explain only a small portion of the variance in grades, suggesting they play a role but don't capture the full picture. Notably, their impact is stronger in lower SES groups compared to higher ones

We visualized the correlations of the climate variables, with the bootstrapping 95% CI for each SES group and compared the resulting patterns.



We can see from the results that, violence prevention strongly correlates with higher scores across all groups. Surprisingly, teacher care is linked to lower scores, especially in higher SES schools perhaps due to less personal attention in high performing environments. Student satisfaction shows a consistent positive effect, while actual violence has little impact except in low SES schools. These are correlations, not causal effects, and missing non-climatic factors may explain some findings (Simpson's paradox).

In conclusion, our question was which factors influence student performance. We found that while climate variables explain only a small part—compared to factors not included in our data, such as study hours, they still have a stronger impact on students from lower socioeconomic backgrounds then higher ones.

4. Limitations and Future Work

One major limitation of our analysis is the large number of missing values across many survey questions in more climate categories such as teacher-student relationship. Additionally, we lacked data for the same schools over multiple years, preventing us from analyzing trends over time.

If we had more time, we would collect more datasets related to more academic skills like intervention of parents, hours of study and more to see the bigger picture.

5. Appendix

- Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). *A Review of School Climate Research.* Review of Educational Research. https://doi.org/10.3102/0034654313483907
- OECD (2019). PISA 2018 Results (Volume III): *What School Life Means for Students' Lives*. https://www.oecd.org/education/pisa-2018-results-volume-iii-acd78851-en.htm
- Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2017). *A Research Synthesis of the Associations Between Socioeconomic Background, School Climate, and Academic Achievement*. Review of Educational Research, https://doi.org/10.3102/0034654316669821
- 2. Taob center report, Feb 2020, https://www.taubcenter.org.il/wp-content/uploads/2020/12/achievementsandgapsheb.pdf