Analysis of the Effect of Student Demographics and Parental Involvement on Student's Academic Performance in the School District of Beloit

By Maclaryn Leonard and Satirtha Saha Protya

Purpose of the study

This case study aims to analyze data collected through the parents and student body of the School District of Beloit to determine potential policies that may positively impact the district's academic performance. By analyzing the information available and suggesting ways of implementing systems to not only these target demographics but the entire school system as a whole, there is a greater capability to strengthen the community and increase test scores. By creating a situation in which all students can receive the best possible education, the school will grow and the community as a whole. Ensuring that the school, families, and students are all involved in the education process there is more availability for growth both academically and within the surrounding community. From our analysis, we were able to observe that the key variables that impacted test scoring were English proficiency as well as parental involvement along with demographic influences such as race, sex, and income. Based on this analysis we recommend the following policies. Firstly, we recommend that the school district find a way to implement a way to increase parental involvement through a variety of ways. Secondly, We recommend that the school provide additional extensive English proficiency schooling to the entirety of the student body.

The study

This analysis was conducted within the School District of Beloit (SDB) surrounding their third-grade MAP testing which takes place three times throughout the year. Through this testing, the SDB has been suffering from poor academic performance and our goal is to provide insight into ways to improve these test scores. This case study allows us as analysts to interpret potential issues and bring about improvement through recommended policy implementations. The questions we have decided on for this study are based on the analysis of parental involvement and English proficiency leading the poor academic performance within this school district. The first question we would like to focus on is, How does the amount of parental involvement have an impact on a student's testing capabilities? By asking this question we can analyze the number of students at each level of parental involvement to determine if that is a concern and focus on its impact on the school district's academic performance. Our secondary question inquires about how English proficiency within differing racial groups has an impact on reading test scores. By answering this question we are able to determine the effects of English proficiency as well as racial differences between reading test scores. Through the information we received from Professor Diep Phan that had been collected from federal information, the school district of Beloit as well as during parent-teacher conferences using a survey, we can interpret and analyze our focus questions through the use of descriptive statistics through a summary report as well as demographic data visualizations which allowed us to determine our key variables required for these questions along with Ordinary Least Squares (OLS) Regression to model the relationship between independent variables (predictors) and dependent variables (academic performance measures).

Data Exploration and Cleaning

The dataset initially provided to us contained 106 columns and data from 221 students, including 2 duplicate student IDs. Key variables included demographic information such as sex, race, and English proficiency, as well as school details (both old and new), health, and special education status. The dataset also featured five years of time-series academic performance data, covering math and reading test scores up to 3rd grade, along with responses from a survey conducted during a parent-teacher conference by Dr. Diep Phan's research team.

For the data cleaning process, we began by removing unnecessary columns that were not relevant to our hypothesis or model building. We then developed a script to summarize each column, allowing us to understand the diversity within the dataset. Based on this analysis, we removed values with very low frequencies, as they contributed little to the overall analysis. A script was also created to convert string values into numerical ones. Additionally, we derived new variables such as Per_Person_Budget, calculated by dividing a household's yearly income by the number of household members, and Parental_Involvement_Score, which was the sum of numerical values assigned to responses from the parental involvement survey questions (6.3-6.8). Lastly, students were excluded from the dataset if their parents left more than one parental involvement question unanswered. Some of the unanswered/missing data were filled with the maiden value for that variable manually if possible. After completing these steps, we were left with key variables including Student ID, Sex, Race, English Proficiency, Average Math and Reading Test Scores for the 2018-2019 school year, Total Score for that year, Per Person Budget, Parental Involvement Score, and the related survey questions with their responses. This refined dataset contained 177 unique student entries for all these variables.

As noted earlier, we created a script to generate summary reports for any variables of interest. From this, we found that our final dataset included 96 female-identified and 81 male-identified students. The majority of the student population (47%) identified as white. While a large portion of the students were highly proficient in English, the dataset also included students from all six other English proficiency categories. The average math test score for the dataset was 189.1 and the average reading test score was 171.58. Additionally, the Parental Involvement Score displayed considerable diversity, as shown in the graph below. These basic analyses helped us select and finalize our key variables.

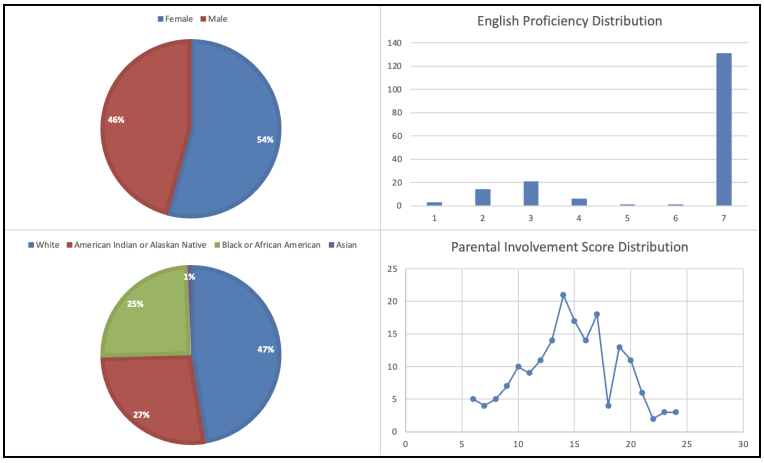

Figure: Basic visualizations showing the diversity among the Sex, Race, English Proficiency, and Parental Involvement Score

Analysis and Findings

To begin our analysis, we explored potential correlations between reading and math test scores. One of our policy recommendations centers around improving English proficiency to boost reading test scores, which we hypothesized would also positively impact math scores. This hypothesis stems from the idea that up until 3rd grade, math is more language-based and descriptive, meaning higher language proficiency could lead to better overall academic performance.

We conducted a simple linear regression to test the relationship between reading and math test scores. Our regression results showed a moderate positive relationship between the two, with an R-squared value of 0.296, indicating that nearly 30% of the variance in math scores could be explained by reading scores. The p-value for this model was highly significant ($p < 0.001$), supporting our hypothesis that students who perform well in reading tend to do well in math as well.

Next, we constructed a multiple regression model to investigate the impact of various factors only on reading test scores. The model included predictors such as sex, race, English proficiency, per-person budget, and parental involvement in different survey questions. The R-squared value for this model was 0.276, meaning that about 27.6% of the variability in reading test scores could be explained by the included variables. We also calculated a Total Parental Involvement Score by summing the scores from all the individual survey questions and tested it as a single predictor. However, the R-squared value for this model was notably low, and we realized that not all the questions could be weighted equally in terms of their impact. Combining them into a single score proved to be somewhat misleading. Therefore, we decided to include each question as a separate variable in our final model to better capture the distinct influence of different aspects of parental involvement.

Key findings from the regression analysis include:

- Sex: Male students were predicted to score lower on reading tests compared to female students, with a statistically significant coefficient (-15.86, $p = 0.033$).
- English Proficiency: This variable had a significant positive effect on reading scores (coefficient = 7.71, $p = 0.003$), affirming our expectation that higher proficiency in English correlates with better academic performance.
- Parental Involvement (Q6_8): This variable, which measures how strictly parents enforce rules regarding homework, was also a significant predictor (coefficient = 15.58, $p = 0.001$). Students whose parents were more involved in monitoring homework tended to perform better in reading.
- Other Parental Involvement (Q6_3, Q6_6): These variables, which measure the frequency of parents attending school events and helping their child with homework, showed positive coefficients. However, the high p-values associated with these variables reduce the statistical confidence in their significance.
- Other Variables: The per-person budget had a small but noticeable effect on reading scores, with a coefficient of 0.0005 per dollar increase. While this impact seems minimal on a per-dollar basis, it could translate into a more substantial effect with larger budget increases, such as $1,000. Additionally, race variables showed some influence on reading scores, with white students

performing better in this model. However, due to high p-values, these findings were not statistically significant.

Key Insights:

1. English Proficiency: As anticipated, students with higher English proficiency scores performed significantly better in reading tests, supporting the idea that language skills play a critical role in academic success.
2. Parental Involvement: The parental involvement score, particularly regarding how parents regulate homework time, was a significant predictor of reading performance, highlighting the importance of family engagement in a child's education.
3. Sex Differences: Female students outperformed male students in reading tests, which might indicate the need for targeted interventions to support male students in this area.

Overall, these findings suggest that improving English proficiency and increasing parental involvement could be key levers for enhancing student academic outcomes.

Policy Recommendations

Through our analysis of parental involvement in relation to test scoring, we were able to determine that an increased amount of parental involvement led to an increase in test scoring. Due to this analysis, we would recommend that the School District of Beloit implement a policy in which all parents regardless of involvement score, although they may want to press this matter more urgently yet privately with parents of lower interaction scores, have access to volunteer whether it be after school or implemented into the classroom curriculum that they can involve themselves with their child's schooling and the school as a whole. By implementing this policy and increasing the parental involvement with their student the analysis has shown that there will be greater academic performance.

Our secondary policy that we would recommend would be an inclusive English proficiency curriculum in which all students regardless of race or English proficiency will take to help the school district to increase performance in these reading-based state tests. Through our analysis of this focus question, we were able to acknowledge that many individuals in the district regardless of racial identity and level of English proficiency are struggling in the reading tests. By assisting in the education of English to students of all identities we are able to increase the test results as well as ensure that no student feels isolated or targeted due to their test performance, race, or English proficiency level. Through our regression, we are able to determine that increasing overall English proficiency throughout all racial identities will have a positive influence on test scores.

Some ethical concerns that we came across while interpreting our data to create our recommended policies include discrimination, anonymity, and issues revolving around biases. As addressed above we recommend that these policies be implemented throughout the entirety of the School District of Beloit to ensure that no student or family feels discriminated against, targeted, or threatened due to being singled out due to a certain demographic influence. By using student ID numbers we are able to ensure that the data used remains anonymous and is unable to be used to target certain individuals or their families. Finally, we wanted to ensure that it was understood that there is bias within this data set surrounding the idea that the survey data was collected during parent-teacher conferences in which mainly parents with a

high level of parental involvement are present. We also recognize that there are several key variables that could influence student success, and by excluding these variables, we introduced omitted variable bias.

Conclusion

In conclusion, we have identified that English proficiency and parental involvement are key variables that have been shown to have an effect on the School District of Beloit's test scores during this time. Through our analysis, we believe that these key factors could have a favorable impact on testing within this school district if addressed appropriately and effectively. To assist in the increase of test scores within this student population we are proposing a parental involvement policy as well as a policy regarding increased education surrounding English proficiency without the discrimination or targeting of a group of any specific groups or individuals. In the future, it would benefit the school district to continue to collect testing scores as well as familial information in order to ensure that proper individuals are able to get the help they need to ensure their students' success. By including information such as the parental English proficiency as well as their health and schooling information we believe that there could be further avenues to explore regarding any genetic impacts on academic achievement.

N.B:- We acknowledge that AI was used in this project, primarily to refine, rewrite, or correct our code and address any errors.