# A Statistical Analysis of Parent Involvement and Child Academic Achievement by School Level and Geographic Location


## Zachary Allen[1], Luke Androski[1], Anthony Hernandez[1], and Finian Hanley[1]

[1]*Department of Statistics, University of Florida, Gainesvile, FL*



### Abstract

The relationship between parent involvement and child academic performance has been extensively studied in the field of education, but the impact of parent involvement has primarily been examined in specific case studies which focus on a singular age or region demographic. This report aims to investigate such relationships at the elementary, middle, and high school level, as well as different demographic communities such as urban, suburban, and rural areas across four major U.S regions using various statistical modeling techniques. Through a comprehensive review of existing literature and an analysis of relevant survey data, a final model was created that suggests the *Parenting* and *Communicating* types of parent involvement have a significantly positive impact on student grades, but are less present in rural and Midwest areas. These results allow recommendations to be made for parents, educators, and policymakers on how to promote successful parent involvement in their areas.

## Background and Significance

The role of parental involvement has been studied and proven in research to drastically improve many aspects of a child's academic life. For instance, Cheung (2019) found parent involvement in child learning to be associated with the child's perception of their relationships with their teachers, which in turn improved their academic performance over time. On a similar note, an overwhelming majority of students want parents to take a higher role in schooling and communication with their school (Epstein 2009). Certain meta-analysis studies, in an attempt to determine the efficacy of parent involvement programs and types, have found parent involvement to differ in certain communities (Jeynes 2012) or specific school levels (Hill and Tyson 2009), however, we seek to analyze how different types of parent involvement impact the academic performance of students at a multitude of school levels, geographic locations, and demographic communities to provide focused recommendations for parents, educators, and policymakers.

Academic performance and parental involvement can take on various definitions, and for the purposes of this study academic performance is measured by the parent's reporting of the student grades and parental involvement will be divided into six categories as defined by Epstein (2009). The six categories being measured are *Parenting*, how home environments are established to support children as students; *Communicating*, what forms of school-to-home and home-to-school communications about school programs and child progress exist; *Volunteering*, the recruiting and organization of parent help and support; *Learning at Home*, how ideas are provided to families on how to help students at home with homework and other related activities; *Decision Making*, how families are included as participants in school decisions; and *Collaborating with Community*, how resources and services from the community are integrated to strengthen school programs, family practices, and student development.

## Methods

The Parent and Family Involvement in Education (PFI) survey collects data about students enrolled in K-12 schools and contains roughly 150 questions about school choice and various aspects of parental involvement where the parents or guardians are the survey respondents. A subset of the most recent edition of this survey (2019) that uses 75 of the most relevant questions and only contains responses from non-homeschooled children will be the sole data source being analyzed. The response variable will be SEGRADES in all cases and represents the parent's reporting of their child's current grades (1 = Mostly A's, 2 = Mostly B's, 3 = Mostly C's, 4 = Mostly D's, 5 = School does not use this grading system).

Out of all respondents of the survey, any that contained invalid or missing responses for any variables of interest or did not use traditional A-F letter grades (SEGRADES = 5) were removed. The remaining data is split into three sections based on if the child is in Elementary, Middle, or High school to allow for individual models to be fit and compared at each level.

For curating variables, 16 binary variables corresponding to selected Yes/No questions from the 2019 PFI survey were used and grouped as indicators for five of the six types of parent involvement (Table A1). The number of "Yes" responses to each of the questions in a given group were used as a sum-variable (`parenSum`, `commSum`, `volSum`, `dmSum`, and `collabSum`) and predictor in the final model. The Learning at Home (LH) type of parent involvement is not considered as its corresponding questions

contain continuous variables that disrupt model interpretations.

Following this, three correlation matrices were used at each level to determine if multicollinearity exists within the predictors and a Cumulative Logit Model (CLM) of Proportional Odds was fit to predict student grades using the predictors mentioned previously (B.1). Parent involvement types found to have the most positive impact on improving student grades were examined for further analysis on how they vary among different geographical regions and community types. The geographical region where the child attends school is represented by CENREG (1 = Northwest, 2 = South, 3 = Midwest, 4 = West) and the community type of the zip code where the school is located is represented by ZIPLOCL (11-13 = City, 21-23 = Suburb, 31-33 = Town, 41-43 = Rural). To examine the statistical relationships of these variables of interest, the CLM of Proportional Odds will be re-fit on four subsets of data corresponding to each of the four CENREG regions regardless of school level. Any significantly positive predictors that are also noticeably different from other regions will be used in pairwise chi-squared tests (B.2) between the predictors, ZIPLOCL, and CENREG. The residuals from these tests will also be plotted.

## Results

The parameter correlation matrices (Table A2) at all school levels show at most moderate correlation between predictors ($< 0.42$) and the estimated and exponentiated coefficients of the CLM per level are shown below in Table 1. For every school level, the *Parenting* and *Communication* types of parent involvement have significantly positive estimates while *Volunteering* and *Collaborating with Community* types have significantly negative estimates. The estimate for *Decision-Making* is not significant at any reasonable significance level.

**Table 1.** Summary Output of CLM on School Level Data

| School Level | Variable | Estimate | exp(Estimate) | p-value |
|---|---|---|---|---|
| Elementary | parenSum | 0.31205 | 1.366225 | 1.50e-09 |
| | commSum | 0.37387 | 1.453344 | 1.16e-13 |
| | volSum | -0.47882 | 0.619512 | <2e-16 |
| | dmSum | 0.06721 | 1.069524 | 0.207 |
| | collabSum | -0.16106 | 0.851243 | 1.42e-13 |
| Middle | parenSum | 0.56025 | 1.7511165 | <2e-16 |
| | commSum | 0.42594 | 1.5310248 | <2e-16 |
| | volSum | -0.48858 | 0.6134993 | <2e-16 |
| | dmSum | -0.07200 | 0.9305285 | 0.226 |
| | collabSum | -0.10859 | 0.8971025 | 5.68e-07 |
| High | parenSum | 0.657659 | 1.9302688 | <2e-16 |
| | commSum | 0.227761 | 1.2557854 | 3.24e-15 |
| | volSum | -0.465061 | 0.6280969 | <2e-16 |
| | dmSum | -0.003942 | 0.9960661 | 0.932 |
| | collabSum | -0.132991 | 0.8754730 | 8.81e-15 |

The refit CLM of Proportional Odds split on census region rather than school level (Table A3) have results consistent with the original CLM, with the exception of `parenSum` in the Midwest, which has a significantly higher exponentiated coefficient (2.0841) compared to all other regions. Thus, pairwise chi-squared tests were conducted between `parenSum`, ZIPLOCL, and CENREG, all of which showed statistical dependence between the variables (p-value $< 2.2 \times 10^{-16}$). The chi-squared residual plots (Fig. A1) show a significantly higher amount of children in rural (Fig. A1(i)) and Midwest (Fig. A1(iii)) areas with a `parenSum` value of 0 than what would be expected under independence. In addition, there is a significantly greater amount of rural communities in the Midwest compared to all other regions than what would be expected under independence (Fig. A1(ii)).

## Discussion and Conclusions

Parent involvement in early child development has already been proven in previous research to provide immense benefits in different aspects of their life. Based on the results from the CLM, it is clear that the *Parenting* and *Communicating* types of parent involvement have the greatest impact on whether a student has good grades. Specifically, the impact from the *Parenting* type increases from Elementary to High School whereas *Communicating* is generally more effective in earlier years.

The geographical analysis that split the dataset on four U.S regions revealed that in the Midwest, holding all other parental involvement types constant, the *Parenting* type of parental involvement has a more positive impact on student grades compared to all other regions. Further analysis of the residual plots (Fig. A1) show that both rural and Midwest locations have a significantly less amount of *Parenting* compared to large suburbs and cities, and the Midwest has a larger portion of rural areas than what would be expected under independence. In short, the Midwest is mainly rural and the odds of student grades improving when the *Parenting* type is present increase by at least 22.45% compared to other regions, however, the *Parenting* type exists much less in these areas.

While the findings of this project are important there are several limitations that should be considered that may impact the interpretation of these results. The first of which being the use of student grades as a proxy for child success has resulted in estimated coefficients that directly contradict decades of research. For instance, the *Volunteering* and *Collaborating with Community* parent involvement types in our research are shown to negatively impact student grades implying that parents should avoid such involvement with their child to improve performance. While this is true in the context of the problem, it would be irresponsible to do so in practice solely based on these models since there are many confounding variables outside of grades that can influence the overall success of a student in which the *Volunteering* and *Collaborating with Community* types are important. Secondly, the distribution of student grades per school level (Table A4) show that an overwhelming majority of students at all school levels have Mostly A's and B's, making it more difficult to determine the true impact of parent involvement in terms of its ability to improve student grades. Lastly, the student's grades are reported by the parent rather than pulled directly from the school, opening the possibility for misreporting that could effect results. Ideas for future research that would address these limitations involve using multiple data sources in conjunction that measure a variety of student success metrics or conduct a deeper dive on the root causes for parents not being involved in their children's lives. From a technical perspective, resampling techniques could be attempted to balance the distribution of grades within the dataset.

There are two primary recommendations for school districts looking to apply the results of this study. The first of which is ensuring parents are maintaining a positive outlook towards their child's education and establishing home environments that allow the child to succeed by requiring schools to have consistent communication between the parent and school that discuss current parent involvement and the academic status or future of the student. Secondly, extra resources should be allocated by school districts within rural locations as they are performing poor in the aforementioned categories. Specifically in the Midwest, where the region is primarily rural, and the *Parenting* type of parental involvement has a significantly greater effect on student grades compared to other regions but is severely lacking in practice.

## References

Cheung, C. S. 2019. “Parents’ involvement and adolescents’ school adjustment: Teacher-student relationships as a mechanism of change.” *Sch Psychol* 34 (4): 350–362. https://doi.org/10.1037/spq0000288.

Epstein, J. L. 2009. *School, Family, and Community Partnerships: Your Handbook for Action.* Calif: Corwin.

Hill, N. E., and D. F. Tyson. 2009. “Parental involvement in middle school: a meta-analytic assessment of the strategies that promote achievement.” *Dev. Psychol.* 45 (3): 740–763. https://doi.org/10.1037/a0015362.

Jeynes, W. 2012. “A meta-analysis of the efficacy of different types of parental involvement programs for urban students.” *Urban Educ. (Beverly Hills. Calif.)* 47 (4): 706–742. https://doi.org/10.1177/0042085912445643.

## Appendix A

**Table A1.** Variables used and their corresponding Parent Involvement Type and Survey Question

| Variable Name | P.I. Type | Question |
|---|---|---|
| SEFUTUREX | Parenting (P) | Do you expect your child to earn a Bachelor's Degree or higher? |
| SCCHOICE | Parenting (P) | Did you feel that you had a choice in what school this child attends? |
| FSATCNFN | Communicating (C) | This school year, has any adult in this child's household gone to a regularly scheduled parent-teacher conference with this child's teacher? |
| FSCOUNSLR | Communicating (C) | This school year, has any adult in this child's household met with a guidance counselor in person? |
| FSMTNG | Communicating (C) | This school year, has any adult in this child's household attended a general school meeting? |
| FSVOL | Volunteering (V) | This school year, has any adult in this child's household served as a volunteer in this child's classroom or elsewhere in the school? |
| FSFUNDRS | Volunteering (V) | This school year, has any adult in this child's household participated in fundraising for the school? |
| FSPTMTNG | Decision-Making (DM) | This school year, has any adult in this child's household attended a meeting of the parent-teacher organization or association? |
| FSCOMMTE | Decision-Making (DM) | This school year, has any adult in this child's household served on a school committee? |
| FSSPORTX | Collab. w Commun. (CC) | This school year, has any adult in this child's household attended a school or class event? |
| FOCRAFTS | Collab. w Commun. (CC) | In the past week, has anyone in your family done activities like arts and crafts? |
| FOGAMES | Collab. w Commun. (CC) | In the past week, has anyone in your family played board games or did puzzles with him or her? |
| FOBUILDX | Collab. w Commun. (CC) | In the past week, has anyone in your family worked on a project like building/fixing something? |
| FOSPORT | Collab. w Commun. (CC) | In the past week, has anyone in your family played sports, active games, or exercised together? |
| FOLIBRAYX | Collab. w Commun. (CC) | In the past month, has anyone in your family visited a library with this child? |
| FOBOOKSTX | Collab. w Commun. (CC) | In the past month, has anyone in your family visited a bookstore with this child? |

**Table A2.** Correlation Matrices At Each School Level and All Levels Combined

| | Elementary | | | | | Middle | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Predictors** | parenSum | commSum | volSum | dmSum | collabSum | parenSum | commSum | volSum | dmSum | collabSum |
| parenSum | 1 | 0 | -0.02 | 0.01 | -0.04 | 1 | 0.03 | 0.01 | 0.02 | -0.04 |
| commSum | 0 | 1 | 0.25 | 0.29 | 0.19 | 0.03 | 1 | 0.26 | 0.32 | 0.22 |
| volSum | -0.02 | 0.25 | 1 | 0.36 | 0.31 | 0.01 | 0.26 | 1 | 0.4 | 0.3 |
| dmSum | 0.01 | 0.29 | 0.36 | 1 | 0.24 | 0.02 | 0.32 | 0.4 | 1 | 0.23 |
| collabSum | -0.04 | 0.19 | 0.31 | 0.24 | 1 | -0.04 | 0.22 | 0.3 | 0.23 | 1 |
| | **High** | | | | | **All Subsets Combined** | | | | |
| parenSum | 1 | 0 | -0.07 | -0.03 | -0.04 | 1 | 0 | -0.05 | -0.01 | -0.07 |
| commSum | 0 | 1 | 0.3 | 0.39 | 0.29 | 0 | 1 | 0.29 | 0.35 | 0.27 |
| volSum | -0.07 | 0.3 | 1 | 0.42 | 0.29 | -0.05 | 0.29 | 1 | 0.41 | 0.35 |
| dmSum | -0.03 | 0.39 | 0.42 | 1 | 0.24 | -0.01 | 0.35 | 0.41 | 1 | 0.26 |
| collabSum | -0.04 | 0.29 | 0.29 | 0.24 | 1 | -0.07 | 0.27 | 0.35 | 0.26 | 1 |

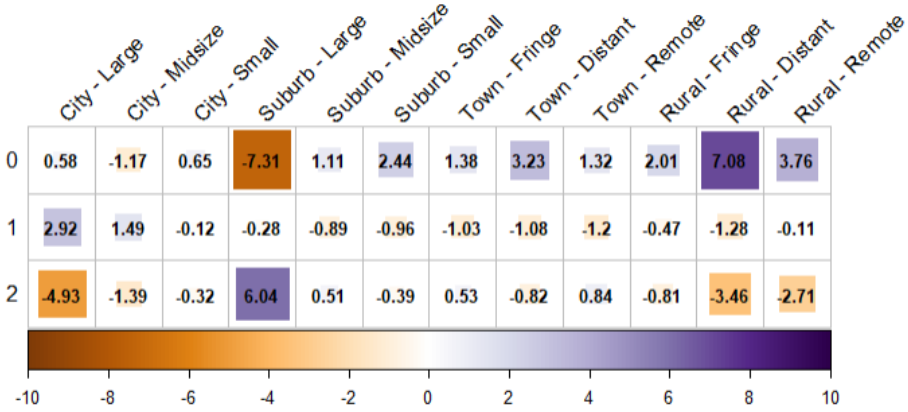
**Table A3.** Exponentiated Coefficient Estimates of CLM on Region Data

| **Region** | parenSum | commSum | volSum | dmSum | collabSum |
|---|---|---|---|---|---|
| Northeast | 1.7019 | 1.2201 | 0.6432 | 1.1187 | 0.8990 |
| South | 1.5900 | 1.4291 | 0.5927 | 0.9668 | 0.8493 |
| Midwest | 2.0841 | 1.2500 | 0.5963 | 1.0249 | 0.8719 |
| West | 1.6203 | 1.4487 | 0.6667 | 0.9349 | 0.8586 |

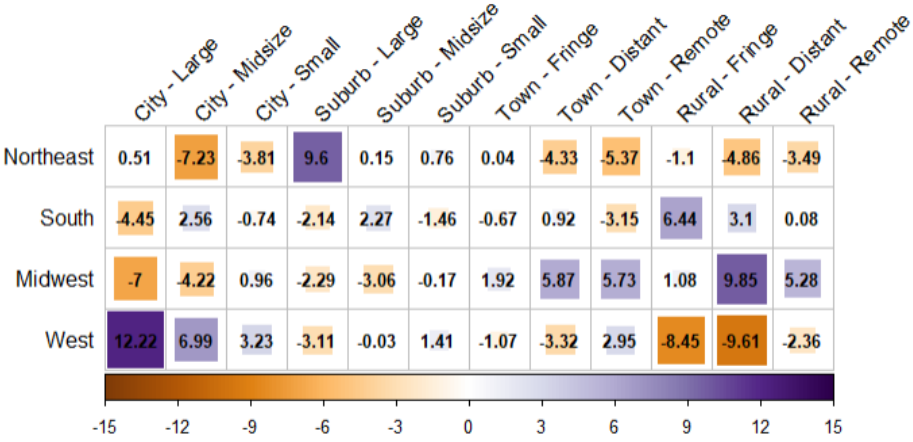
**Table A4.** Distribution of Respone Variable (SEGRADES) at Each School Level

| **Reported Grades** | % of Elementary School Students | % of Middle School Students | % of High School Students | % of All Students |
|---|---|---|---|---|
| Mostly A's | 63.022 | 56.191 | 51.856 | 56.451 |
| Mostly B's | 29.664 | 32.290 | 33.975 | 32.198 |
| Mostly C's | 6.091 | 9.853 | 11.845 | 9.540 |
| Mostly D's or Lower | 1.223 | 1.666 | 2.324 | 1.811 |

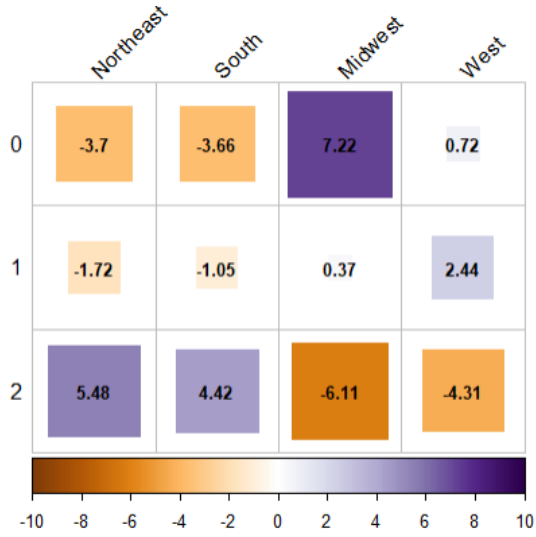
**Figure A1.** Pairwise Chi-squared Residual Plots of ZIPLOCL, CENREG, and `parenSum` with Test p-values




(i) `parenSum` and ZIPLOCL (p-value $< 2.2 \times 10^{-16}$)




(ii) CENREG and ZIPLOCL (p-value $< 2.2 \times 10^{-16}$)




(iii) `parenSum` and CENREG (p-value $< 2.2 \times 10^{-16}$)

## Appendix B

### 0.1 Cumulative Logit Model of Proportional Odds

The Cumulative Logit Model (CLM) of Proportional Odds has the following definition given the context of the problem.

$$\log\left(\frac{P(Y \leq j)}{1 - P(Y \leq j)}\right) = \alpha_j + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$$

where $x_1 = parenSum, x_2 = commSum, x_3 = volSum, x_4 = dmSum, x_5 = collabSum$ and $j \in [1, 2, 3]$ corresponds to a response category (SEGRADES). This model determines the log odds that a certain child's predicted grade level category is less than or equal to $j$ (grades at or above a certain letter). For instance, $\log(odds(P \leq 2))$ would be the log odds that a given student's reported grades are "Mostly B's or Above". Therefore, the odds of a student with grades at or above a certain level multiply by $e^{\beta_i}$ for each 1-unit increase in $x_i$ holding all other variables constant. Parameters with significantly positive estimated coefficients indicate that their corresponding parent involvement type has a positive impact on student grades. The property of proportional odds is assumed to be held even if the non-proportional odds model shows statistical significance (p-value = 0.01619) to reduce the number of estimated coefficients from 18 (15 $\beta$'s and 3 $\alpha$'s) to 8 (5 $\beta$'s and 3 $\alpha$'s) making interpretations and comparisons between school level models simpler.

### 0.2 Chi-squared Tests of Independence

The chi-square test for independence has the following null and alternative hypothesis:

$H_0$ : There is independence between the two variables being tested

$H_a$ : There is dependence between the two variables being tested

The expected value of each cell ($E_{ij}$), residual of each cell ($r_{ij}$), and overall test statistic ($X^2$) are as such:

$$E_{ij} = \frac{(\text{row total})(\text{column total})}{\text{total sample size}} \qquad r_{ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}}} \qquad X^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where $O_{ij}$ is the observed value in the $i$th row and $j$th column and the degrees of freedom are equal to $df = (r - 1)(c - 1)$. The null hypothesis is rejected if $P(\chi^2_{df} \geq X^2) < \alpha$. Significantly positive residuals indicate more observed values than what would be expected under the null hypothesis and the opposite is true for significantly negative residuals. Pairwise residual plots are shown in Figure A1.