HARVARD
GRADUATE SCHOOL OF EDUCATION

Multilevel Modelling: How Are Socioeconomic Status and Per-Pupil Spending

Associated With Variation in Elementary Student Achievement Across U.S. School Districts

Sylvia Li

S052: Intermediate and Advanced Statistical Methods for Applied Education Research

Introduction

Decades of research have shown that SES is one of the strongest predictors of academic achievement, with children from wealthier backgrounds benefiting from a range of advantages both inside and outside the classroom (Duncan & Murnane, 2011). Meanwhile, debates about school funding raise important questions: does spending more actually help students succeed, or does it depend on where and how that money is used (Baker, 2016)?

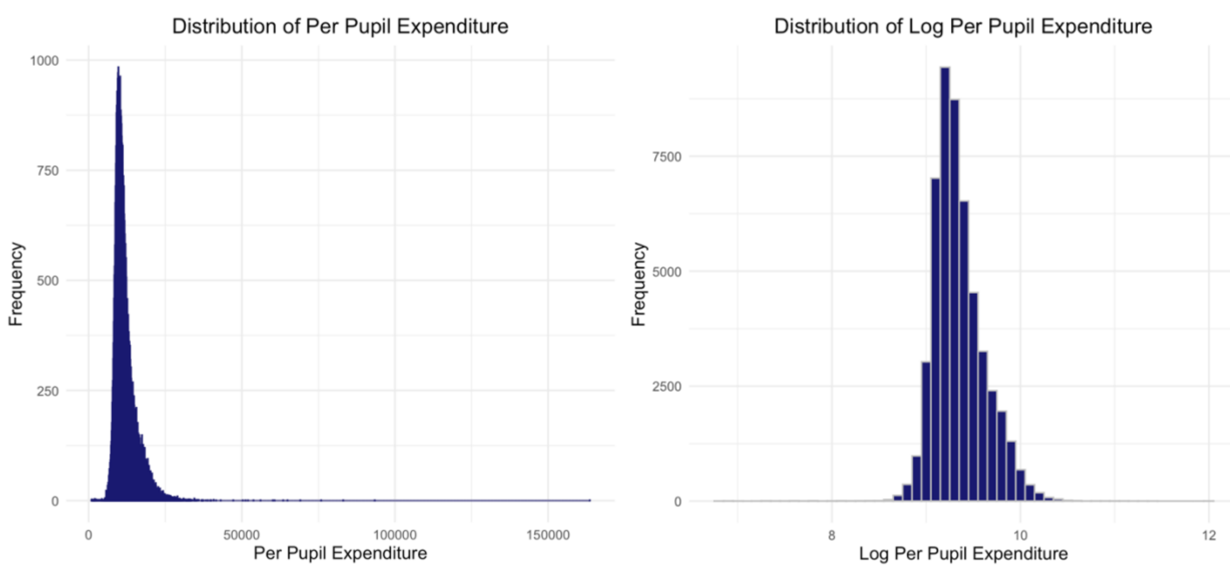
In this project, I explore how district-level SES and per-pupil school spending relate to student achievement in U.S, drawing on data from the Stanford Education Data Archive (SEDA) (Stanford University) and the Common Core of Data (CCD) (The Institute of Education Sciences) for per-pupil spending (NCES, 2015–2019) for student achievement and socioeconomic indicators. Rather than trying to establish cause and effect, my aim is to understand patterns: are wealthier, better-funded districts consistently performing better? And does spending make more of a difference in districts with lower SES? Using multilevel models, I look at both changes within districts over time and differences between them. This approach helps unpack how local conditions may interact to shape student outcomes, reflecting my continued interest in the structures behind educational opportunity.

Exploratory Data Analysis (EDA)

For my analysis, I brought together data from several sources to build a clean dataset at the district-year level. I started with the Standford Education Data Archive (SEDA) data, using th *geodist_long_cs* format, where it contains geographic school district achievement estimates, long by grade-year-subject. First, I calculated the weighted average test score across grades 3 to 8 for each district and year to simplify the structure. Then, I turned to NCES fiscal data and calculated per-pupil spending, by dividing total expenditures by student enrollment. After that, I merged in socioeconomic status indicators ($sesall$) from SEDA's covariate file. With all three key variables: test scores, spending, and SES, aligning by district and year, I cleaned the dataset by removing any missing or zero values. Finally, I log-transformed the Per Pupil Spending to prepare it for modeling. This gave me a solid, ready-to-use dataset that captures the financial and social context behind student achievement across U.S. school districts.

Figure 1

Per Pupil Spending before (left) and after (right) log-transformation



Note. Per pupil spending is constructed by total expenditure (TOTALEXP) / Fall Membership (V33) in the Fiscal Common Core of Data, since it does not contain the raw data of Per Pupil Spending at district levels.

In Figure 1, The left panel shows a highly skewed distribution of raw per pupil expenditure, with most districts spending under $25,000 but a long right tail of higher values. After log transformation (right panel), the distribution becomes approximately normal, which is more suitable for regression modeling. This improves interpretability and reduces the influence of extreme outliers.

Figure 2

Overall Time Trend Test Score by SES Level

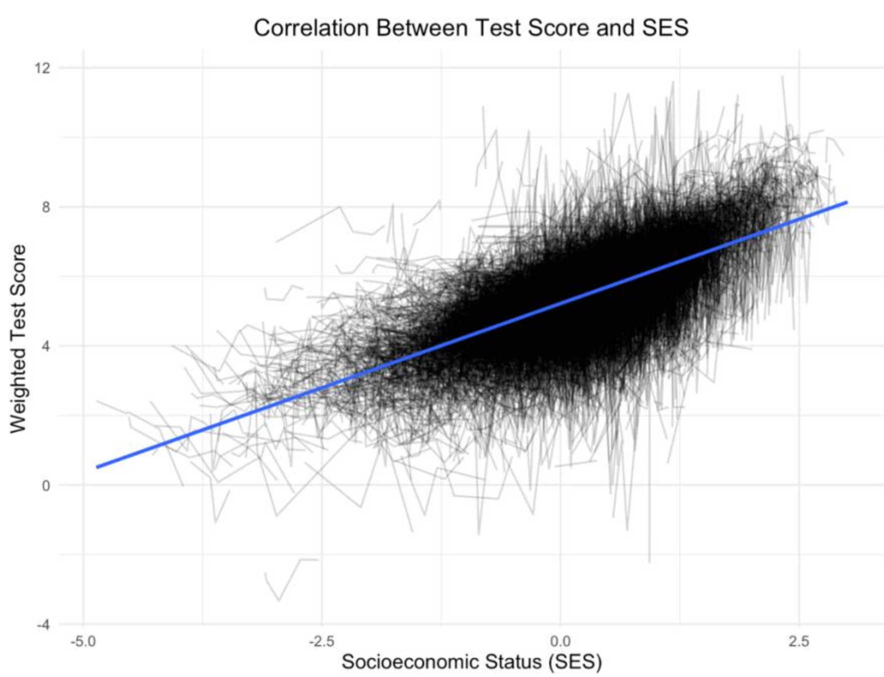
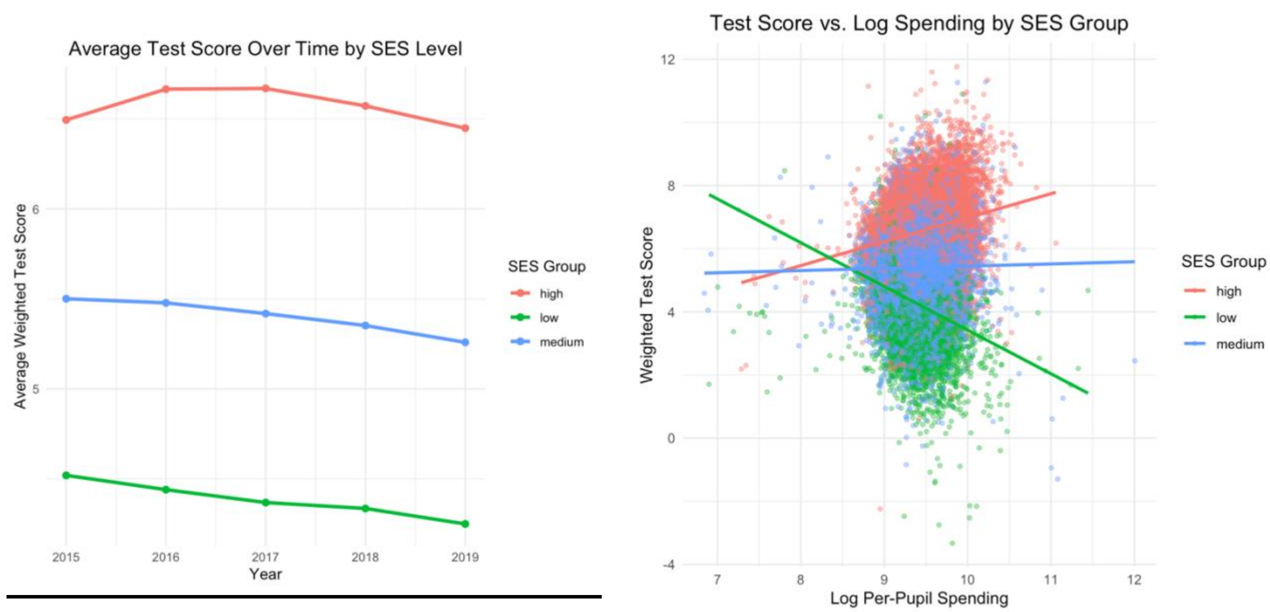


Figure 2 visualizes average test score trends over time (2015–2019) across all the districts, illustrating the overall relationship between district socioeconomic status (SES) and weighted

test scores. T black lines represent a district-year observation repeated observations from the same district over time. The clear upward-sloping blue line represents a fitted linear trend, indicating a strong positive association: as SES increases, so does academic achievement. This pattern underscores one of the central themes of the research question – districts with higher socioeconomic status tend to have significantly better student outcomes, which warrants deeper exploration into how SES interacts with other factors like school funding.

Figure 3

Average Test Score by Time (Left) & Test Score Vs. Per Pupil Expenditure (Right)



In **Figure 3**, the *Left* plot it illustrates average test score trends over time (2015–2019) across three SES groups: low, medium, and high. The SES groupings were created using distribution cutoffs: below the 1st quartile as low, between the 1st and 3rd quartiles as medium, and above the 3rd quartile as high. A persistent gap in achievement can be observed across SES groups, with high-SES districts consistently scoring highest and low-SES districts scoring lowest, showing signs of slight decline over time. The *Right* plot explores the relationship between log per-pupil spending and test scores, stratified by SES level. This more closely resembles a fixed effect, specifically the relationship of log spending and SES. Interestingly, the trend lines suggest that higher spending is positively associated with achievement in high-SES districts, but the association is flat or even negative in lower-SES districts. Together, these figures highlight the need to further investigate not only whether money matters, but for whom and under what socioeconomic conditions. These patterns speak directly to the research question, prompting deeper inquiry into how district-level SES and financial resources interact to shape academic opportunity.

Methodology

I estimated three multilevel linear models with random intercepts, to examine the relationship between district-level socioeconomic status (SES), per-pupil spending, and student achievement. The population models of the student test score on a set of covariates. These models progressively include within- and between-district SES, financial resources, and their interaction, accounting for the nested structure of time within districts using random intercepts at the district level.

<u>Model 1:</u> Adding fixed effect (SES)

$$meanavg_{td} = \beta_0 + \beta_1 SES_{td} + \varepsilon_{td} + \zeta_d$$

$$\zeta_d \sim N(0, \sigma_\zeta^2)\ ;\ \varepsilon_{td} \sim N(0, \sigma_\varepsilon^2)$$

Where:

- SES_{td}: time-varying SES (within-district variation)
- $\zeta_d \sim N(0, \sigma_\zeta^2)$: random intercept for district

- $\varepsilon_{td} \sim N(0, \sigma_\varepsilon^2)$: residual

This model estimates the year-to-year changes in district SES on student achievement, controlling for unobserved differences between districts.

Model 2: Adding Mundlak decomposition mean of SES and per pupil expenditure

$$meanavg_{td} = \beta_0 + \beta_1 SES_{td} + \beta_2 \overline{SES_d} + \beta_3 \log(spending_{td}) + \varepsilon_{td} + \zeta_d$$

$$\zeta_d \sim N\left(0, \sigma_\zeta^2\right);\ \varepsilon_{td} \sim N(0, \sigma_\varepsilon^2)$$

Model 2 estimates the separate association of within- and between-district SES and spending, by adding the additional variables using a Mundlak Contextual Effect (Mundlak, 1978), where:

- SES_{td}: **Time-varying SES**, capturing within-district changes over time
- $\overline{SES_d}$: **District-mean SES**, capturing between-district socioeconomic context.

The Mundlak specification separates within- and between-group variation by including the group mean of a time-varying predictor as an additional covariate (Mundlak, 1978). This allows to distinguish how changes in SES within a district over time (SES_{td}) differ from differences between districts' long-term average SES ($\overline{SES_d}$), estimating how both changes in a district's SES and its long-term socioeconomic context relate to test scores, while adjusting for financial resources available per student.

Model 3: Adding cross-level interaction

$$meanavg_{td} = \beta_0 + \beta_1 \overline{SES_d} + \beta_2 SES_{td} + \beta_3 \log\,(spending_{td})$$
$$+\beta_4 (\overline{SES_d} \times spending_{td}) + \varepsilon_{td} + \zeta_d$$

$$\zeta_d \sim N(0, \sigma_\zeta^2)\ ;\ \varepsilon_{td} \sim N(0, \sigma_\varepsilon^2)$$

Model 3 builds upon Model 2 by adding an interaction between SES and per-pupil spending. This cross-level interaction tests whether the relationship between a district's SES and its student achievement is moderated by the level of financial resources available. For example, it allows us to explore whether the benefits of increased SES are more pronounced in high-spending districts, or conversely, whether spending has a greater relationship in lower-SES settings. By modeling this interaction, we could investigate whether spending can compensate for or amplify SES-related inequalities in educational outcomes.

Results

Multilevel Model Results: SES and Achievement

	Test Score (1)		**Test Score (2)**		**Test Score (3)**	
Predictors	*Estimates*	*std. Error*	*Estimates*	*std. Error*	*Estimates*	*std. Error*
(Intercept)	5.27 ***	0.01	4.41 ***	0.20	5.34 ***	0.21
sesall	0.68 ***	0.01	-0.00	0.01	0.02	0.01
ses_mean			1.00 ***	0.02	-2.04 ***	0.20
log(spending)			0.08 ***	0.02	-0.02	0.02
ses_mean × log(spending)					0.32 ***	0.02
Random Effects						
σ^2	0.27		0.26		0.26	
τ_{00}	0.76 sedalea		0.68 sedalea		0.67 sedalea	
ICC	0.74		0.73		0.72	
N	10757 sedalea		10757 sedalea		10757 sedalea	
Observations	51152		51152		51152	
Marginal R^2 / Conditional R^2	0.267 / 0.807		0.455 / 0.851		0.460 / 0.850	

* *p<0.05* ** *p<0.01* *** *p<0.001*

In Model 1, the coefficient for $sesall$ is **0.68** (SE = 0.01, p < .001)**,** indicating a statistically significant positive association between year-to-year changes in district SES and student achievement. This means that when a district's SES improves from one year to the next, its average test scores tend to rise by 0.68. The high ICC of 0.74 suggests that the 74% of the variation in scores exists between districts rather than within each district over time. However, sesall (SES_{td}) estimate becomes 0 in Model 2, while the coefficient for ses_mean ($\overline{SES_d}$) is **1.00** (SE = 0.02, p < .001)**.** This indicates that long-term differences in SES across districts when they become time invariant, are more strongly associated with achievement. The continued high ICC **(0.73)** confirms that district-level differences remain the dominant source of variance.

Model 3 shows the interaction between ses_mean and log_spending is positive and statistically significant **(**0.32, SE = 0.02, p < .001**),** while the main associations of both ses_mean (**–2.04**) and log_spending (**–0.02**) become negative. This reflects a moderating relationship: the association of average SES on achievement grows stronger in districts that spend more. Substantively, this means high-SES districts benefit more from increased spending, while low-SES districts see little or no gain from spending alone. The interaction reveals that the advantage of SES is amplified in well-funded contexts**,** and that the relationship between SES and achievement is not uniform, which depends on the financial resources available.

To have a more specific interpretation of the ICC, it decreases slightly from 0.74 in Model 1 to 0.73 in Model 2 and 0.72 in Model 3**,** suggesting that adding between-district SES and its interaction with spending explains a small portion of the between-district variance. Intuitively, this means that while districts still differ substantially from one another in test scores (high between-group variance), those differences are becoming slightly more predictable based on SES and spending. At the same time, the high ICC also implies that students within the same district tend to perform similarly, reinforcing the idea that local context, rather than individual fluctuation, is a strong driver of achievement.

Conclusion

This study began with a personal interest in understanding the structural factors that shape educational opportunity across school districts. The findings indicate that long-term socioeconomic status is a strong predictor of student achievement, while the impact of school spending is most pronounced in districts with higher socioeconomic advantage. These results suggest that funding alone may not be sufficient to close achievement gaps, its effectiveness appears to depend on the broader context in which it is allocated.

There are, however, important limitations to this analysis. First, it is important to note that this study is observational and descriptive in nature; it does not establish a causal relationship between SES, school spending, and achievement. The associations observed reflect patterns across districts but may be confounded by unmeasured variables or reverse causality. To strengthen causal inference in future research, approaches such as instrumental variables, regression discontinuity, or natural experiments could be employed to better isolate the effects of spending or SES. Incorporating longitudinal designs or exploiting policy changes could also help identify mechanisms and causal pathways more precisely.

Second, State-level factors, that potentially indicating the distinctive policy environments and funding structures for each state, were not included as a Level 3 variable in the multilevel models in this analysis. As a result, the innate differences between states remain unaccounted for, potentially limiting the precision of the estimates.

Future research could improve upon this work by incorporating state-level variation, applying methods to address missing data, and exploring how specific spending categories, such as teacher salaries or student support services that affects students' education opportunity. This project has deepened my interest in educational inequality and highlighted the need for more nuanced investigations into how resources and social context interact to influence student success.

References

Baker, Bruce. "Does Money Matter in Education? Second Edition." *Shanker Institute*, 20 Jan. 2016, www.shankerinstitute.org/resource/does-money-matter-second-edition.

Duncan, Greg J, and Richard J Murnane. "Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances." 2012.

Mundlak, Yair. "On the Pooling of Time Series and Cross Section Data." *Econometrica*, vol. 46, no. 1, Jan. 1978, p. 69, https://doi.org/10.2307/1913646. Accessed 9 May 2025.

Reardon, Sean F. "The Widening Academic Achievement Gap between the Rich and the Poor." *Inequality in the 21st Century*, 15 May 2018, pp. 177–189, https://doi.org/10.4324/9780429499821-33. Accessed 9 May 2025.

The Institute of Education Sciences. "Common Core of Data (CCD)." *Nces.ed.gov*, nces.ed.gov/ccd/files.asp#Fiscal:1.

University, Stanford. "The Educational Opportunity Project at Stanford | SEDA Data Downloads." *The Educational Opportunity Project at Stanford University*, 2019, edopportunity.org/opportunity/data/downloads/. Accessed 9 May 2025.

Appendix A

R Code for Data Cleaning

```
# install.packages("sas7bdat")

library(sas7bdat)

library(haven)

library(lme4)

library(dplyr)

library(tidyr)

library(ggplot2)

library(tidyverse)

library(stargazer)

seda <- read.csv('/Users/le/Desktop/2025 Spring/052 final
project/052_rproject/SEDA/seda_geodist_long_gcs_5.0.csv')

cov <- read.csv('/Users/le/Desktop/2025 Spring/052 final
project/052_rproject/SEDA/seda_cov_geodist_long_5.0.csv')

seda <- seda %>% filter(year >= 2015)

cov <- cov %>% filter(year >= 2015)

#Extract useful columns

seda_cleaned <- seda %>% select(sedalea, stateabb, sedaleaname, year, grade, gcs_mn_all,
tot_asmt_all)



#1. weighted average of test score across grades



cov %>% summarize(n = n_distinct(sedalea))

seda %>% summarise(n = n_distinct(sedalea))
```

```
 # Calculate weighted average for each district-year

seda_cleaned <- seda_cleaned %>%

  group_by(sedalea, sedaleaname, year) %>%

  summarise(weighted_avg_testscore = sum(gcs_mn_all * tot_asmt_all) /
sum(tot_asmt_all), .groups = "drop")

#  # Join the result back to the original seda dataframe

# seda <- seda %>%

#   left_join(district_year_avg, by = c("sedalea", "year"))

#2. expenditure per pupil

#2019

exp_19 <- read_sas('/Users/le/Desktop/2025 Spring/052 final project/052_rproject/18-
19/sdf19_2a_sas.sas7bdat')

summary(exp_19$V33)

summary(exp_19$TCURELSC)

exp_19 <- exp_19 %>%

  select(LEAID, V33, TCURELSC) %>%

  filter(V33 >= 0) %>%

  filter(TCURELSC >= 0) %>%

  mutate(exp_per_pupil = TCURELSC / V33 )



#2018

exp_18 <- read_sas('/Users/le/Desktop/2025 Spring/052 final project/052_rproject/17-
18/sdf18_1a.sas7bdat')

summary(exp_18$V33)

summary(exp_18$TCURELSC)
```

```
exp_18 <- exp_18 %>%

  select(LEAID, V33, TCURELSC) %>%

  filter(V33 >= 0) %>%

  filter(TCURELSC >= 0) %>%

  mutate(exp_per_pupil = TCURELSC / V33 )


#2017

exp_17 <- read_sas('/Users/le/Desktop/2025 Spring/052 final project/052_rproject/16-17/sdf17_1a.sas7bdat')

summary(exp_17$V33)

summary(exp_17$TCURELSC)


exp_17 <- exp_17 %>%

  select(LEAID, V33, TCURELSC) %>%

  filter(V33 >= 0) %>%

  filter(TCURELSC >= 0) %>%

  mutate(exp_per_pupil = TCURELSC / V33 )


#2016

exp_16 <- read_sas('/Users/le/Desktop/2025 Spring/052 final project/052_rproject/15-16/sdf16_1a.sas7bdat')

summary(exp_16$V33)

summary(exp_16$TCURELSC)
```

```
exp_16 <- exp_16 %>%
  select(LEAID, V33, TCURELSC) %>%
  filter(V33 >= 0) %>%
  filter(TCURELSC >= 0) %>%
  mutate(exp_per_pupil = TCURELSC / V33 )


#2015
exp_15 <- read_sas('/Users/le/Desktop/2025 Spring/052 final project/052_rproject/14-15/sdf15_1a.sas7bdat')
summary(exp_15$V33)
summary(exp_15$TCURELSC)
exp_15 <- exp_15 %>%
  select(LEAID, V33, TCURELSC) %>%
  filter(V33 >= 0) %>%
  filter(TCURELSC >= 0) %>%
  mutate(exp_per_pupil = TCURELSC / V33 )


#merge them into seda by year
exp_15 <- exp_15 %>% mutate(year = 2015)
exp_16 <- exp_16 %>% mutate(year = 2016)
exp_17 <- exp_17 %>% mutate(year = 2017)
exp_18 <- exp_18 %>% mutate(year = 2018)
exp_19 <- exp_19 %>% mutate(year = 2019)
```

```
exp <- rbind(exp_15, exp_16, exp_17, exp_18, exp_19)


exp <- exp %>%
  select(LEAID, year, exp_per_pupil) %>%
  rename(sedalea = LEAID)
seda$sedalea <- factor(seda$sedalea)
exp$sedalea <- factor(exp$sedalea)
cov$sedalea <- factor(cov$sedalea)
seda_cleaned$sedalea <- factor(seda_cleaned$sedalea)
seda %>% distinct(sedalea) %>% count()
exp %>% distinct(sedalea) %>% count()
cov %>% distinct(sedalea) %>% count()
seda_cleaned %>% distinct(sedalea) %>% count()
exp <- exp %>%
  mutate(sedalea = sub("^0+", "", sedalea))
seda_cleaned <- seda_cleaned %>% left_join(exp, by = c("sedalea", "year"))
#3. join covariates
#collapse the covariates to the same level as seda
cov <- cov %>%
  select(sedalea, year, sesall) %>% filter(year >= 2015)
cov <- cov %>%
  group_by(sedalea, year) %>%
  summarise(sesall = first(sesall), .groups = "drop")
```

```
#join ses

seda_cleaned <- seda_cleaned %>% left_join(cov, by = c("sedalea", "year"), relationship =
"many-to-many")

seda_cleaned <- seda_cleaned %>% rename(weighted_score = weighted_avg_testscore )

#taking a mundlak mean

# Decompose SES into between and within components

seda_cleaned <- seda_cleaned %>%

  group_by(sedalea) %>%

  mutate(

    ses_mean = mean(sesall),   # Between-district effect

    # Within-district (raw SES_it)

  ) %>%

  ungroup()

# 4. check the missing values

seda_cleaned %>% filter(is.na(weighted_score)) %>% count()

summary(seda_cleaned$exp_per_pupil)

summary(seda_cleaned$sesall)

summary(seda_cleaned$ses_mean)

summary(seda_cleaned$weighted_score)




seda_cleaned <- seda_cleaned %>%

  filter(!is.na(weighted_score),

         !is.na(sesall),
```

```
        !is.na(ses_mean),

        !is.na(exp_per_pupil),

        exp_per_pupil > 0)

seda_cleaned <- seda_cleaned %>%

  mutate(log_spending = log(exp_per_pupil))

write.csv(seda_cleaned, '/Users/le/Desktop/2025 Spring/052 final
project/052_rproject/seda_cleaned.csv', row.names = FALSE)
```

Appendix B

R Code for Exploratory Data Analysis (EDA)

```
library(ggplot2)

#Figure 1: the distribution of spending

ggplot(seda_avg, aes(x = exp_per_pupil)) +

  geom_histogram(binwidth = 100, fill = "purple", color = "midnightblue", alpha = 0.7) +

  labs(title = "Distribution of Per Pupil Expenditure", x = "Per Pupil Expenditure", y =
"Frequency") +

  theme_minimal()+

  theme(plot.title = element_text(hjust = 0.5,  size = 14),

        axis.title.x = element_text(size = 12),

        axis.title.y = element_text(size = 12))


#the distribution of log spending

ggplot(seda_avg, aes(x = log_spending)) +

  geom_histogram(binwidth = 0.1, fill = "midnightblue", color = "grey") +
```

```
  labs(title = "Distribution of Log Per Pupil Expenditure", x = "Log Per Pupil Expenditure", y = 
"Frequency") +

  theme_minimal()  +

  theme(plot.title = element_text(hjust = 0.5, size = 14),

                        axis.title.x = element_text(size = 12),

                        axis.title.y = element_text(size = 12))


#Figure 2:  relationship between test score and SES

ggplot(seda_avg, aes(x = sesall, y = weighted_score, group = sedalea)) +

  geom_line(alpha = 0.2) +  # Line for each district

  geom_smooth(se = FALSE, method = "lm", aes(group = 1)) +  # Overall trend line

  labs(title = "Correlation Between Test Score and SES",

       x = "Socioeconomic Status (SES)",

       y = "Weighted Test Score") +

  theme_minimal() +

  theme(

    plot.title = element_text(hjust = 0.5, size = 14),

    axis.title.x = element_text(size = 12),

    axis.title.y = element_text(size = 12),

    legend.position = "none"  # Hide legend if there are many districts

  )


#Figure 3. ahievement by ses level
```

```r
# Left Plot

# Group SES into low, medium, high (already done)

seda_cleaned <- seda_cleaned %>%

  mutate(ses_group = case_when(

    sesall <= -0.3 ~ "low",

    sesall > -0.3 & sesall <= 0.82 ~ "medium",

    sesall > 0.82 ~ "high"

  ))

# Summarize average scores by year and SES group

summary_data <- seda_cleaned %>%

  group_by(year, ses_group) %>%

  summarize(avg_score = mean(weighted_score, na.rm = TRUE), .groups = "drop")


ggplot(summary_data, aes(x = year, y = avg_score, color = ses_group, group = ses_group)) +

  geom_line(linewidth = 1.2) +

  geom_point(size = 2) +

  labs(title = "Average Test Score Over Time by SES Level",

       x = "Year",

       y = "Average Weighted Test Score",

       color = "SES Group") +

  theme_minimal() +

  theme(plot.title = element_text(hjust = 0.5, size = 14))
```

```
#Right Plot. test score vs log spending by SES


# Create SES group using defined cutoffs
seda_cleaned <- seda_cleaned %>%
  mutate(ses_group = case_when(
    sesall <= -0.3 ~ "low",
    sesall > -0.3 & sesall <= 0.82 ~ "medium",
    sesall > 0.82 ~ "high"
  ))
# Scatterplot with linear trend lines by SES group
ggplot(seda_cleaned, aes(x = log_spending, y = weighted_score, color = ses_group)) +
  geom_point(alpha = 0.4, size = 1) +
  geom_smooth(method = "lm", se = FALSE, linewidth = 1) +
  labs(title = "Test Score vs. Log Spending by SES Group",
       x = "Log Expenditure Per-Pupil",
       y = "Weighted Test Score",
       color = "SES Group") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, size = 14))
```

Appendix C

R Code for Modelling

```
install.packages("sjPlot")
```

```
library(sjPlot)

seda_cleaned <- read.csv('/Users/le/Desktop/2025 Spring/052 final
project/052_rproject/seda_cleaned.csv')

#Model 1: This model estimates the proportion of test score variance that lies between districts
(for ICC).

mod1 <- lmer(weighted_score ~ sesall + (1 | sedalea), data = seda_cleaned, REML = FALSE)

summary(mod1)

#Model 2: Add main effect: This estimates separate relationships of SES (within and between)
and spending.

mod2 <- lmer(weighted_score ~ sesall + ses_mean + log_spending + (1 | sedalea), data =
seda_cleaned)

summary(mod2)

#Model 3: Test whether the relationship of SES depends on funding, i.e., does funding amplify
or weaken the SES association on test score?

mod3 <- lmer(weighted_score ~ sesall + ses_mean + log_spending + ses_mean:log_spending +
(1 | sedalea), data = seda_cleaned)

summary(mod3)

tab_model(mod1, mod2, mod3,

    show.ci = FALSE,        # Hides confidence intervals

    show.se = TRUE,         # Shows standard errors

    show.p = FALSE,         # Hides raw p-values

    show.stat = FALSE,      # Keeps test statistics (optional)

    show.icc = TRUE,        # Shows ICC

    show.re.var = TRUE,     # Shows random effects

    show.ngroups = TRUE,    # Shows number of districts

    dv.labels = c("Test Score(1)", "Test Score(2)", "Test Score(3)"),
```

```
title = "Multilevel Model Results: SES and Achievement",
p.style = "stars")       # Replaces p-values with asterisks
```