THE IMPACT OF RACIAL DIVERSITY ON SCHOOL PERFORMANCE AND

EARLY CHILDHOOD DEVELOPMENT: A CASUAL INFERENCE APPROACH

Long Nguyen

University of South Florida

ISM6642: Advanced Data Science

Prof. Shivendu Shivendu

# Executive Summary

This study aims to examine the impact of racial diversity in schools on early academic performance and children’s development, a topic of significant societal attention, specifically regarding the current US political landscape. The project utilized data from the Early Childhood Longitudinal Studies (ECLS) kindergarten cohort of 2010-11 conducted by the US Department of Education and the National Center for Educational Statistics (NCES). The primary method of study was staggered Difference-in-Difference (DiD) analysis with Fixed Effects, which examines the causal effect of racial diversity, represented by the “Percent of Non-white Students at School”, on IRT-based performance scores. As a result, the findings indicate a positive and gradual impact of attending a diverse school on children’s performance over the semesters. Understanding this relationship, policymakers and educational institutions can strategize DEI initiatives more effectively, while the public can have a more comprehensive understanding of the role of diversity in academia, ultimately helping facilitate the environment where children can thrive and reach their full potential.

# Contents

[Executive Summary 2](#_Toc197638725)

[Contents 3](#_Toc197638726)

[Problem Significance 4](#_Toc197638727)

[Data Source & Preparation 5](#_Toc197638728)

[1. ECLS-K:2011 Kindergarten – Fifth Grade Dataset 5](#_Toc197638729)

[2. Data Preparation 5](#_Toc197638730)

[2.1. Variable Selection 5](#_Toc197638731)

[2.2. Working With the rceth Variables 6](#_Toc197638732)

[2.3. De-flattening the Dataset 6](#_Toc197638733)

[Hypothesis 7](#_Toc197638734)

[Descriptive Analysis 7](#_Toc197638735)

[1. Summary Statistics 7](#_Toc197638736)

[2. Explore Children Performance Throughout the Semesters 9](#_Toc197638737)

[Difference-in-Differences (DID) Analysis and Interpretation 9](#_Toc197638738)

[1. DID and Staggered DID 9](#_Toc197638739)

[2. Fixed Effects Model 11](#_Toc197638740)

[3. Interpretation 12](#_Toc197638741)

[Robustness Checks / Parallel Trends Assumption 13](#_Toc197638742)

[Actionable Insights 14](#_Toc197638743)

[Limitations 15](#_Toc197638744)

# Problem Significance

In recent years, especially around the 2024 US election, the political landscape within the United States has escalated to a great level of tension. There have been ongoing heated conversations over economic policies, social issues, and the role of government. Among different domains being discussed, diversity, equity, and inclusivity (DEI) receives major attention from the broader society. Many claim DEI policies promote the progression of the country toward a just society, in which every individual feels a sense of belonging and understanding. On the other hand, some believe that DEI programs unfairly disadvantage certain groups and hinder the development of society by prioritizing one’s identity over their merits.

Across domains, DEI in the US education system remains a critical controversy. There have been attempts from both points of view to either maintain or remove such initiatives in schools. In fact, the US Department of Education has taken certain actions to eliminate DEI at the beginning of 2025. From the White House government website, President Donald J. Trump has signed an Executive Order to restrict DEI initiatives and refocus school discipline policies on objective behavior. "The Order requires a report to the President that includes an analysis of DEI-based school discipline and its consequences, measures to ensure that federal funds do not support racially preferential policies, including through nonprofit organizations, and proposing model discipline policies rooted in American values.” (The US Government, 2025)

The strong pushback from the federal government has created uncertainty and forced educational institutions to re-evaluate their programs. However, whether these decisions will yield positive outcomes for the education system is still a question to be answered. Many educators and organizations are still actively seeking compliant and effective ways to foster an equitable learning environment for all students. (Ray, 2025)

The objective of this project is to study the significance of DEI on academic performance and children's development. It is essential for policymakers and education institutions to gain a comprehensive understanding of the subject matter to support their decision-making process. If the principles of DEI demonstrate a positive impact on children, specifically at an early age, continued commitment to DEI programs might be vastly beneficial long-term. This would ask the government and some state legislatures to reconsider their anti-DEI policies. Furthermore, this study seeks to address disparities and biases regarding DEI in education in today’s society, hopefully identifying opportunities to shape an environment where children can flourish academically, socially, and emotionally.

# Data Source & Preparation

## 1. ECLS-K:2011 Kindergarten – Fifth Grade Dataset

This project will analyze the data from the Early Childhood Longitudinal Studies (ECLS) Program, conducted by the US Department of Education and the National Center for Educational Statistics (NCES, n.d.).

The ECLS program examined child development, school readiness, and early school experiences. Moreover, it provided extensive data to examine how various factors: family, school, and community, relate to children’s development and academic performance. Each child record contains data from the child assessments and child questionnaires, the various respondents associated with the child, weights and imputation flags, and administrative variables. This is helpful as the goal is to draw a causal inference between diversity and performance in schools, if they were causally related.

This project looked into the kindergarten class of 2010-11 cohort, which records data from 18,174 participating children between fall 2010 and spring 2016 semesters. The Fall 2014 and Fall 2015 semesters were not recorded. The study followed the same students from their kindergarten year to fifth grade, allowing time-series analysis. The raw dataset from NCES contains 18,174 rows x 26,061 columns, each of which represents the answer to a survey question throughout the whole study period. Due to the massive amount of data and the flattened nature of the dataset, intensive data cleaning and preprocessing are crucial steps before analyzing the causal relationship.

## 2. Data Preparation

### 2.1. Variable Selection

Among 26,061 recorded variables, the following were selected to study the relationship between racial diversity in school and children’s development:

* *childid*: Children ID numbers.
* *s1\_id, s2\_id*, …: School ID numbers. The study was conducted over 9 semesters, and each *sN\_id* indicates the school the child has gone to.
* *x2krceth, x4rceth*, …: Percentage of non-white children at the school. For the second (*x2*) and fourth (*x4*) semesters, the data was collected as a continuous variable ranging from 0 – 100%. The remaining semesters recorded this information as a discrete variable, with 1 meaning 0 – 25%, up to 4 equaling 76 – 100%. No racial data was collected for the first, third, and fifth semesters.
* *x1rthet, x2rthet, x1mthet*, …: IRT-based scores to measure children's performance in reading (*rthet*), math (*mthet*), and science (*sthet*) across the semesters. IRT is a method for modeling assessment data that enables calculation of overall scores for each skill area, allowing fair comparisons between children regardless of their different test items.

Due to missing values, the data subset, after filtering out unnecessary variables, contained 3,177 observations. This should be sufficient to run the analysis and make causal inferences.

### 2.2. Working With the rceth Variables

These variables are the primary indicators of racial diversity of schools. However, the original dataset has multiple issues with these columns. Firstly, the data was collected inconsistently as *x2krceth* and *x4rceth* were recorded as continuous instead of discrete variables like the rest. Therefore, these 2 variables were converted into discrete data, ranging from 1 – 4, to ensure consistency.

Secondly, there was no data recorded for the first (*x1*), third (*x3*), and fifth semesters (*x5*). To solve this issue, I used the school ID numbers of the corresponding semesters to identify the percentage of non-white students at that school during that semester. It should be noted that most students did not switch schools multiple semesters in a row. Thus, for instance, a student’s *x3rceth* is very likely to be similar to either their *x2krceth* or *x4rceth*, depending on whether their *s3\_id* is the same as *s2\_id* or *s4\_id*. Likewise, x5rceth should be equal to either *x4rceth* or *x6rceth*.

Finally, I personally decided *x1rceth* to always be 1 regardless of the school ID number. This is because the first semester should be the baseline, and the earliest the “treatment” can occur should be the second semester.

Additionally, the "treatment" in this study is when a child has been exposed to a racially diverse educational environment. In other words, if they had spent over 4 semesters at schools with *rceth* equal or greater than 2 (more than 25% of students at the school were non-white), then they belong to the *treated* group. *Treated* is a generated variable to separate the treatment and control groups based on the number of semesters the children had spent at diverse schools.

### 2.3. De-flattening the Dataset

In the initial format, all the information for a single child across multiple semesters is compressed into a single row/record. To support analysis, de-flattening the dataset into a time-series format would make it easier to identify the pattern over time.

The process involved iterating through each semester throughout the cohort. For every iteration, the information of each child in the corresponding semester was extracted and restructured into the new time-series data frame. Specifically, for each semester (ranging from 1 to 9), a new series of row was created containing the *childid*, the corresponding semester number (*sem*), the school identifier (*s\_id*), the reported race/ethnicity (*rceth*), the math IRT score (*m\_irt*), the science IRT score (*s\_irt*), the reading IRT score (*r\_irt*), the treatment/control group binary identifier (*treated*), the start date of the treatment (*treatment\_sem*), the number of semesters past the start date (*event\_time*), and pre/post-treatment binary identifier (*post*). Since the columns in the original dataset were named with the semester identifier, it was easy to select the correct values to match the corresponding semester in each row. An average IRT score was also aggregated for each row based on *s\_irt, r\_irt\_* and *m\_irt*. This serves as the primary target variable, showing the overall performance of each child every semester.

This transformation converted the wide and flattened format, in which each child had one row representing data across multiple time points, into a long format where each row represents a single child’s data at a particular period. The restructured data frame is indexed by the *sem* and *childid* variables, allowing for time-series analysis across the cohort.

# Hypothesis

The primary objective of the study was to examine the longitudinal impact of school racial composition on students’ development, demonstrated through their overall performance in reading, science, and math. Hypothetically, the racial composition of a school, measured by the proportion of students of color, would be significantly associated with the academic performance of the children. Racial diversity in the classroom was expected to cause positive influences, considering the potential benefits and challenges of being exposed to different backgrounds and perspectives. At an early age, this would reinforce their ability to adapt to differences and understand the complexity of racial dynamics within and outside school.

# Descriptive Analysis

## 1. Summary Statistics

Fig 1. Overall descriptive analysis of the data set

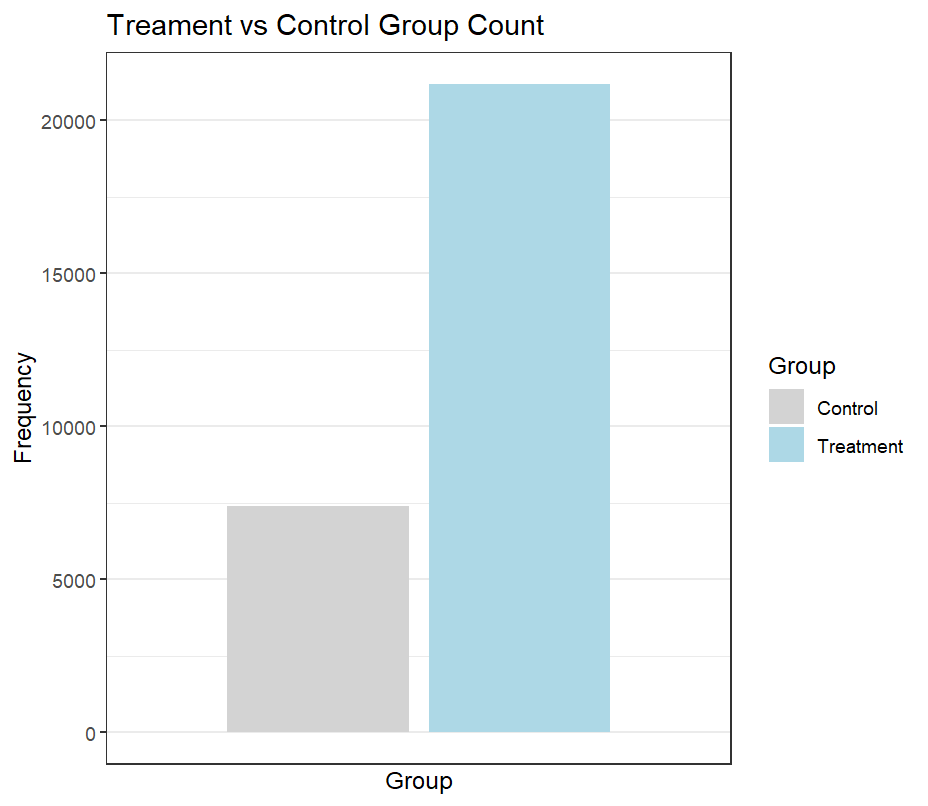
A screenshot of a computer screen

AI-generated content may be incorrect.

Fig 1 describes the values for each of the variables in the data frame:

* As discussed above, *sem* ranges from 1 to 9, showing that the student's progress was tracked over nine semesters.
* *rceth* ranges from 1 to 4, representing four levels of racial diversity at the recorded schools. The average is 2.464, which indicates that the majority of students attended somewhat diverse schools throughout most of the semesters. For this reason, the mean for the *treated* variable is higher than 0.5, proving that most children belong to the treatment group. The frequencies of treatment and control groups can be seen in Fig 2 below.

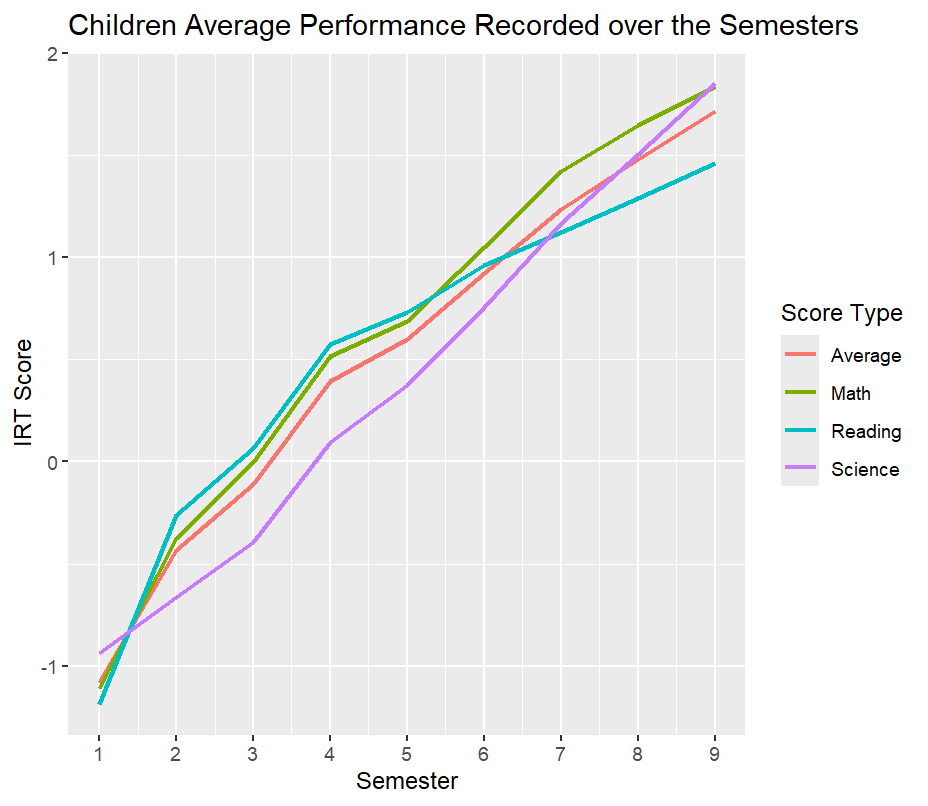
Fig 2. The total count of treatment vs control groups



* *r\_irt, m\_irt, s\_irt*, and *avg\_irt* show the range of the IRT exam scores. The negative values are appropriate as it is common for IRT-based scores, in practice, to be between approximately -3 and 3. The distributions of these scores help understand the relative performance among the students in each category.

## 2. Explore Children Performance Throughout the Semesters

Fig 3. Children's Average IRT-based Scores Over the Semesters



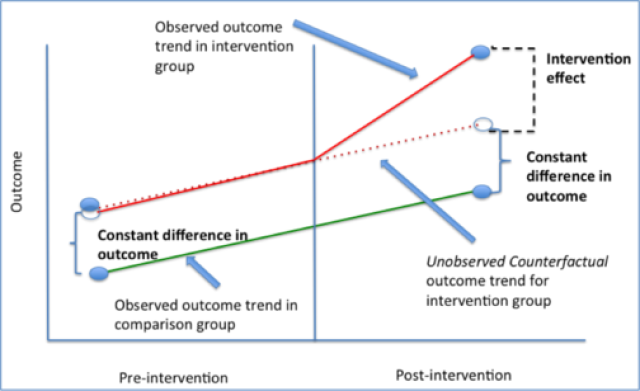
On average, all students started at around -1 in IRT scores across all subjects and began to develop stronger cognitive skills over time. Notably, the trendline for the reading IRT score was steeper than the others in the early semesters. This might be because language, along with reading are foundational skill and the most important focus in early education. Children need to develop linguistic skills before adopting any other subjects in school. Math and science can have more complicated concepts, thus being introduced in later semesters. Apparently, children started to demonstrate stronger improvement in math and science IRT scores in later semesters, surpassing the average reading IRT at the end of the study.

# Difference-in-Differences (DID) Analysis and Interpretation

## 1. DID and Staggered DID

The Difference-in-Differences (DID) analysis is a quasi-experimental design that helps assess the causal effect of an intervention or treatment by comparing the changes in outcomes between the units where the event happened (treatment group) and those where the event did not occur (control group). DID takes advantage of longitudinal data, which our dataset applied, to obtain a counterfactual representing the outcome trend for the treatment group should the event does not occur. (Columbia University Mailman School of Public Health, 2023)

Fig 4. DID Estimation, Graphical Explanation



DID effect is usually implemented as an interaction term between time (indicating the time points before and after the intervention) and binary treated (separating the treatment and control groups) variables in a regression model:

To understand the DID formula, we can input actual values into the model to estimate the difference in change over time between the treatment and control group:

* Change in the control group =
* Change in the treatment group =
* Thus, the difference in changes between the two groups =

Therefore, β3 is the primary coefficient to be examined to understand the causal effect of the intervention.

Furthermore, in some experiments, intervention does not occur at the same time for all observations. Therefore, the model would not have a universal date to separate the pre- and post-intervention periods, as demonstrated in Fig 4. Particularly, in this study, *treatment\_sem*, which represents the first semester that a child attended a diverse school, varies for each individual. For that reason, the staggered DID model was implemented to address the variability in the timing of the treatment, allowing for a more nuanced analysis. Specifically, instead of using the binary *time* variable like in the basic DID model, the variable *event\_time* was created, allowing better analysis of the causal effect over time relative to the *treatment\_sem*.

## 2. Fixed Effects Model

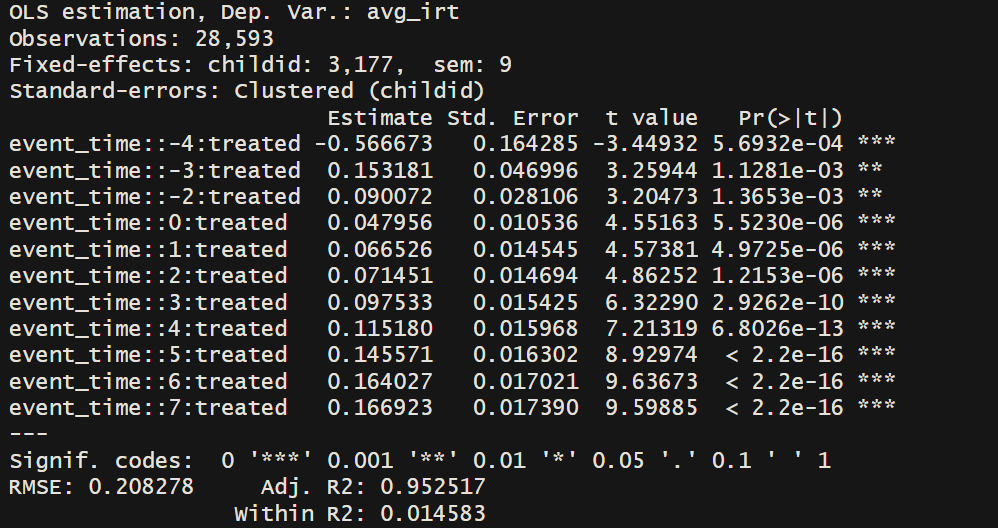
A fixed effects model is a subcategory of regression model. This model is used to control variables that remain constant over time (time-invariant) while isolating the effect of other variables. Regarding this experiment, each child comes with unique traits, natural abilities, or intrinsic motivation. Fixed effects model would help address these factors and focus solely on the changes resulting from attending a more diverse environment.

In this analysis, the average IRT score (*avg\_irt*) is the dependent variable, predicted by the interaction between *event\_time* and treated, representing the causal effect. The fixed effects model also helps control for *childid* and *sem* variable. Including *childid* fixed effects addresses all time-invariant differences between students, as mentioned above. Additionally, including *sem* fixed effects controls for any time-specific effects that are common to all students. Although schools most likely did not follow the same curriculum, each semester should introduce an increase in difficulty. As seen above in Fig 3, students’ IRT-based scores improved over time. Thus, the fixed effects control for both individual-specific and time-specific confounding factors.

Finally, ref = -1 was declared as the baseline for the analysis to allow comparison of treatment and control groups between the post and pre-treatment semesters. Because the earliest *treatment\_sem* is 2, semester 1 (*event\_time = -1*) is the consistent pre-treatment baseline for all observations. The coefficients for the interaction term will then represent the change in *avg\_irt* at *event\_time* t relative to the -1 pre-treatment semester for the treatment group, compared to the control group at the same relative time.

## 3. Interpretation

Fig 5. Fixed Effects Model Output Summary



For each *event\_time,* the coefficients show the estimated difference in *avg\_irt* between treatment and control groups relative to the baseline level. The coefficients at and after the treatment (*event\_time >= 0*) represent the treatment effect. It is notable that the coefficients are statistically significant, positive and consistently increase over time, indicating that attending a diverse school improves avg\_irt and the effect grows larger as time goes by. Specifically, 1 semester after the treatment, the difference in IRT score between the two groups is 0.067 higher than the difference at 1 semester before the treatment. This difference gradually increases to 0.167 at *event\_time* = 7, which is not exactly drastic, but it does present the impact of diversity in school.

On the other hand, before the treatment, the coefficients do not follow a clear pattern. There is a drastic change in the difference between *event\_time* = -4 and *event\_time* = -3, from -0.567 to 0.153, but the coefficient then drops to 0.090 at *event\_time = -2*. Besides, the coefficients for the pre-treatment semesters are all statistically significant and not exactly small compared to the post-treatment coefficients. This might indicate a violation of the parallel trends assumption and needs further checking.

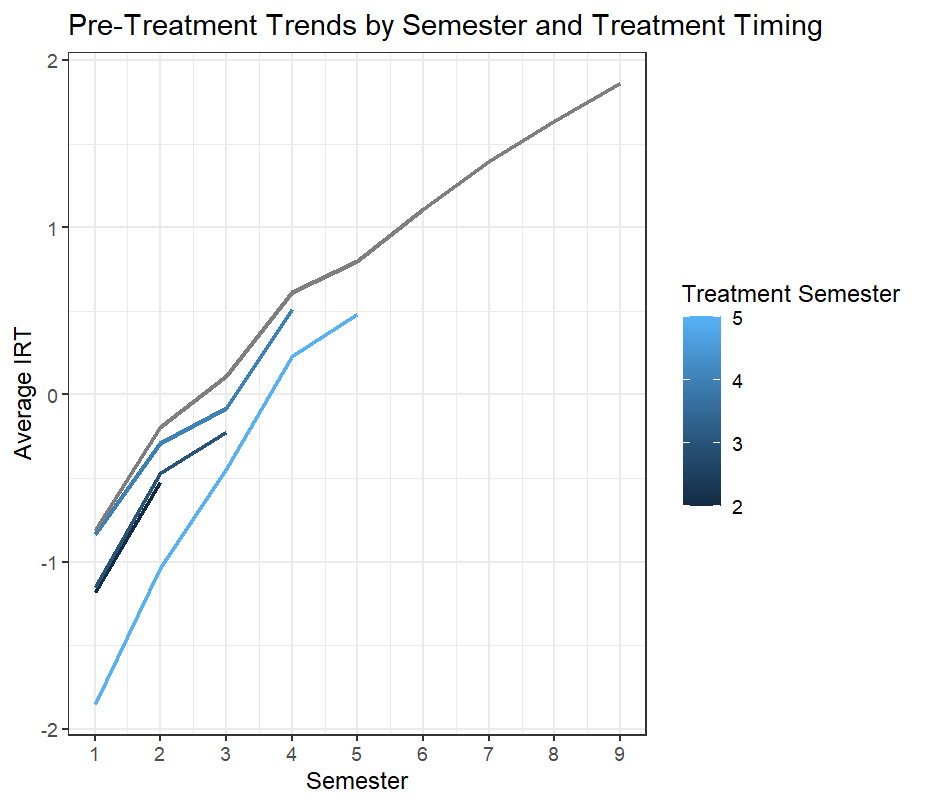
Finally, the fixed effects model returns an adjusted R2 = 0.9525, meaning that the model effectively explains 95.25% of the variance in IRT score. Regarding the within R2 = 0.0145, the treatment effect, once accounting for the fixed effects, only explains a small proportion of the remaining variation in IRT score. This is common for fixed effects models. However, the treatment effect is statistically significant, which shows that being exposed to a racially diverse environment plays a small role in enhancing children’s development and academic performance.

# Robustness Checks / Parallel Trends Assumption

Parallel trend assumption in DID analysis states that the difference between treatment and control groups remains constant over time should the intervention not occur. Violating the parallel trend assumption might lead to biased estimates or incorrect conclusions, which inflate the treatment effect when the intervention actually does not cause the change.

As mentioned above, since there is variation in *treatment\_sem* among the treatment group, there is not a specific point in time to separate the pre- and post-treatment semesters. Therefore, to examine this assumption, the treatment group was split into various subsets depending on their *treatment\_sem*. By doing this, it would be easier to evaluate each subset-specific counterfactual.

Fig 6. IRT Score Trend Before the Intervention (Attending a Diverse School)

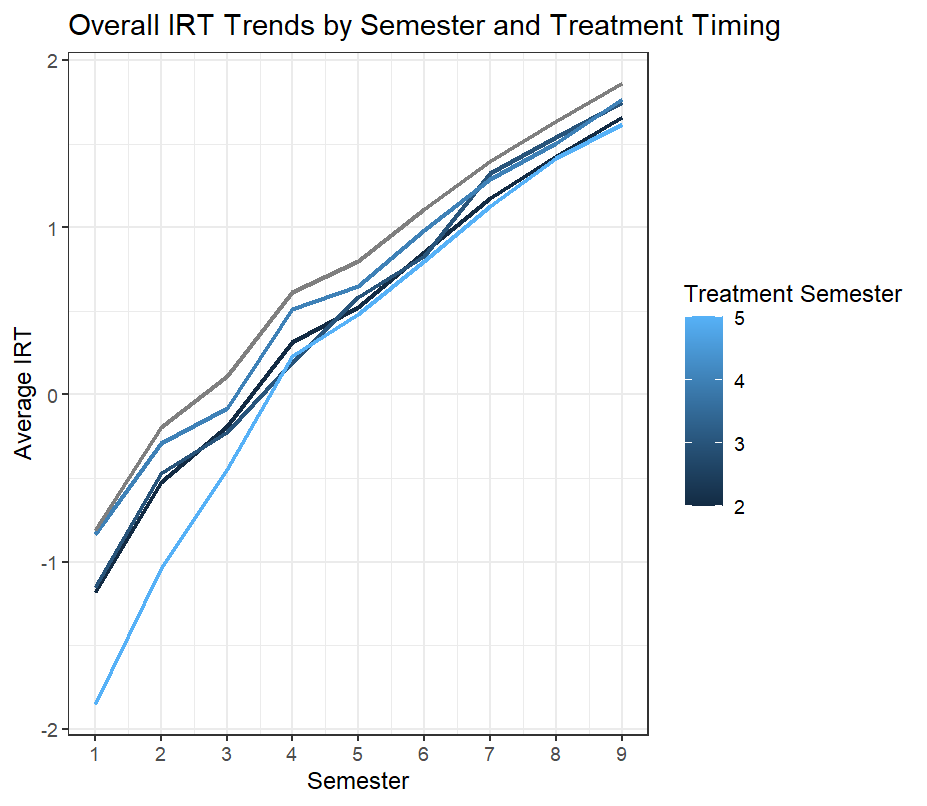


Overall, the groups still meet the parallel trends assumption. To evaluate this assumption, all trend lines were analyzed per period. It could be seen that the subset lines (blue) in the treatment groups mostly have the same slope as the control group pattern (grey). The only point that is slightly concerning is between semesters 2 and 3, in which the patterns of subsets with the *treatment\_sem* = 3 and 4 are not as steep compared to the control group pattern. There is a slight divergence in the line chart, yet nothing seems significant enough to signify the violation of the parallel trends assumption.

# Actionable Insights

In conclusion, the model shows a generally positive and increasing effect of going to a diverse school at a younger age. Although the immediate treatment effect is moderate, the positive impact of racial diversity and inclusivity should be recognized as the growth pattern underscores the potential for long-term academic gains. For that reason, it is recommended that policymakers and school administrators consider implementing DEI policies and programs. The objective of such initiatives is to foster an equitable environment for all students, while allowing opportunities for challenging biases and stereotypes, along with learning different perspectives.

Fig 6. Overall IRT Trends by Semester Across Control and Treatment Groups (Expanded from Fig 5.)



Another important takeaway is that DEI requires long-term effort and commitment to start seeing actual returns on investment. The model's results show a gradual but consistent increase in the treatment effect over the semesters. Hence, it is better to shift away from short-term DEI practices and instead focus on fostering a sustainable and equitable learning environment. DEI should not be just another thing to check off your list, but schools should strive to see, understand, and operate towards equity. According to leadership coach Jamila Dugan, "transformation requires investment in personal and interpersonal development, awareness and creation of shared cultural practices, and the redesign of inequitable systems—all at the same time" (Kohlbecker, 2022)

# Limitations

Firstly, one major limitation of the study lies in how the dependent variable – percentage of non-white children at the school – was utilized. This variable was considered as the sole representation of racial diversity at the schools recorded in the cohort. However, it should be noted that the proportion of non-white students might still be dominated by only one specific racial group. If this were the case, the variable would inaccurately represent diversity in the classroom, leading to inflation in the treatment effect estimate. Furthermore, simply using the racial distribution of schools might not be sufficient to demonstrate the full concept of diversity, which also captures the complex social dynamics and interactions that contribute to the school environment. The treatment in this study was not based on the actual implementation of any DEI initiatives. Thus, it is difficult to precisely investigate the treatment effect.

Nonetheless, the original dataset from NCES consists of multiple variables, including the racial composition of schools and survey questions about school resources and support. Future research could explore these variables to a greater extent to accurately capture the racial diversity in schools. Additionally, there is great potential to expand it over racial diversity and incorporate other factors such as gender or socioeconomic status.

Last but not least, the parallel trends assumption is the most critical assumption for DID analysis. Although the brief analysis in the robustness checks showed that this assumption was met, further testing beyond visualization could be done to ensure that this condition is satisfied. For instance, a pre-treatment model could be utilized to examine the parallel pattern between control and different treatment subgroups.

References

The United States Government. (2025, April 23). *Fact sheet: President Donald J. Trump reinstates commonsense school discipline policies*. The White House. <https://www.whitehouse.gov/fact-sheets/2025/04/fact-sheet-president-donald-j-trump-reinstates-commonsense-school-discipline-policies/>

Ray, V. (2025, April 15). *Area educators affirm commitment to dei init...: Issue 811*. The Commons. <https://www.commonsnews.org/issue/811/811DEI_debate>

*ECLS-K:2011 restricted-use kindergarten-fourth grade data file and electronic codebook*. National Center for Educational Statistics (NCES). (n.d.). <https://nces.ed.gov/use-work/resource-library/data/data-file/ecls-k2011-restricted-use-kindergarten-fourth-grade-data-file-and-electronic-codebook?pubid=2018083>

*Difference-in-difference estimation: Columbia public health*. Columbia University Mailman School of Public Health. (2023, March 13). <https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation>

*Kohlbecker, N. (2022, February 28). Navigating dei in schools: Five crucial considerations. ascd.* [*https://www.ascd.org/blogs/navigating-dei-in-schools-five-crucial-considerations*](https://www.ascd.org/blogs/navigating-dei-in-schools-five-crucial-considerations)