Capstone Progress Report Template

**Stakeholder Names and Roles**

| **Stakeholder** | **Role** |
| --- | --- |
| Russell Carlock | Sponsor: rcarlock@k12albemarle.org |
| Heman Shekari | Team mentor: hs9hd@virginia.edu |
| Stephanie Landas | Team member: sfl7ck@virginia.edu |
| Michelle Wu | Team member: mw3ef@virginia.edu |
| Noah Edwards-Thro | Team member: xjb6yb@virginia.edu |
| Sirish Desai | Team member: skd3nz@virginia.edu |

**Project Title:** Does cultural training of teachers improve student outcomes?

**Abstract**

Research supporting culturally responsive teaching has demonstrated high rates of success in accelerating student learning for students of color with economic hardship. This project will analyze students’ math and reading standardized test scores with classroom teachers going through the culturally responsive teaching and those who have yet to do so. We use student data from the 2023 Albemarle County Public School District to analyze whether the cultural training some teachers received had a statistically significant impact on student academic achievement, specifically, on student reading and math test scores. We will use multivariate linear regression and random forest models, which will highlight the prominent factors for growth of student test scores. The results of this project will inform the Albemarle County Public Schools leadership team on whether or not to implement culturally responsive teaching for all classroom teachers for this next upcoming school year of 2024-2025.

**Outline of the Project**

Research supporting culturally responsive teaching has demonstrated high rates of success in accelerating student learning for students of color with economic hardship. This project will analyze students’ math and reading standardized test scores with classroom teachers going through the culturally responsive teaching and those who haven’t yet. The tools that will be used will be multivariate linear regression and random forest models which highlight the prominent factors for the growth of the students in terms of their test scores. The inputs of the model will be demographic information of the students and teachers and the outputs of the model will be evaluated on the growth of students’ test scores. The results of this project will inform the Albemarle County Public Schools leadership team on whether or not to implement culturally responsive teaching for all classroom teachers for this next upcoming school year of 2024-2025. The assumption is that all of the data for each student has been collected correctly, meaning that any analyses abstracted from it are correctly coded.

**Success Criteria**

The goal as stated by our mentor is to have unbiased knowledge of whether the teacher certification was impactful when it comes to growth of student test scores. Our mentor told us that the most important feature for him in general was how the student grew over the year. He also would like for the findings to be statistically significant. He is also curious about whether the certification could work better for each subject.

However, he also has secondary achievements because he is curious to see what other combinations of variables lead to an increase in growth score. For example, he wants to see if the highest increase in scores could be due to some demographic of the teacher. He is also interested in knowing if the vertical scaled score (achievement scaled across grades and subjects) is influenced by any of the other features.

**Data Assumptions and Limitations**

*Data Limitations*One limitation is due to the Family Educational Rights and Privacy Act (FERPA). As we are working with protected information, no analysis with groups smaller than 10 students/data points can be reported on. One way this has affected the data structure is that students of Asian or multi-racial identities had to be combined into an “other” category for the student race variable.

At the moment, our math dataset does not contain data from fourth and fifth graders, due to unclear records from the schools. We expect this data to be shared in the coming weeks. The mathematics dataset has 258 instances of missing data for the “growth” and “expected achievement” variables and 271 instances of missing data for the “vertical scaled score” variable. The reading dataset has 472 instances of missing data for the “value added score” and “expected achievement” variables, and 66 instances of missing data regarding teacher information (license type, years of experience, highest degree, birth year, gender, and race/ethnicity).

*Data Assumptions*

As we have separate datasets for reading test outcomes and math test outcomes, we are assuming that there is no overlap of students in the dataset. This is in compliance with FERPA as well, as linking a student’s demographic information with both their reading and math scores would allow for identification.

Another assumption is that the data has been correctly aligned. Due to FERPA, we cannot gain access to the raw, confidential information, so we are assuming that the scores are correctly aligned with the student’s demographic information.

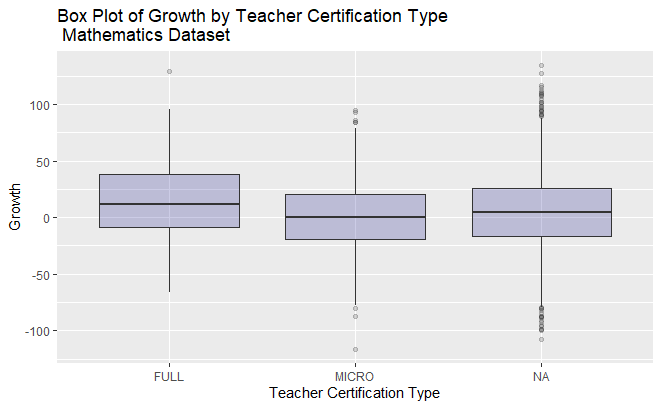
**Summary of Data Processing, Data Aggregation**

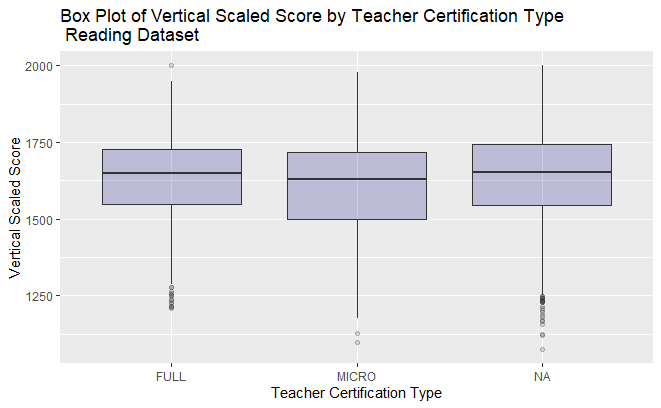
Our data was received in two .csv files (one reading and one math). These files included student demographic information (race, gender, English learner, disabilities, etc.), teacher demographic information (years experience, highest degree, birth year, gender, race, etc.), whether or not a teacher and a principle are certified as culturally responsible educators (and if so, what type of certification), and student results on Virginia SOLs (Standards of Learning) such as growth, expected and actual achievement, and vertical scaled score (achievement scaled across grades and subjects).

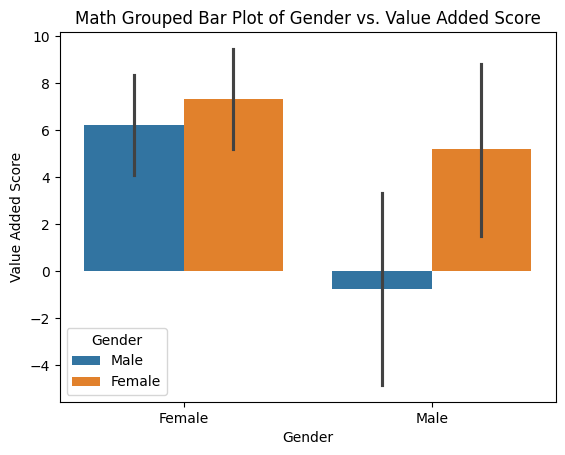
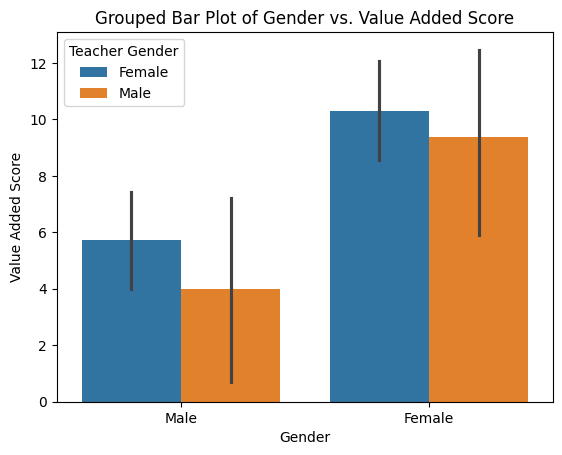
Our data processing only consisted of cleaning the data to make sure the variable types were the same and matching variable names. Being that the goal of our project is to investigate if culturally responsive teaching has any impact on student outcomes, we will attempt to account for the impact of all other variables in relation with whether the teacher is culturally responsive certified. Our potential predictors variables therefore include the teacher and student demographic information (specifically investigating teacher experience/qualification) as well as whether or not a teacher is culturally certified and what type of certification there is. However,

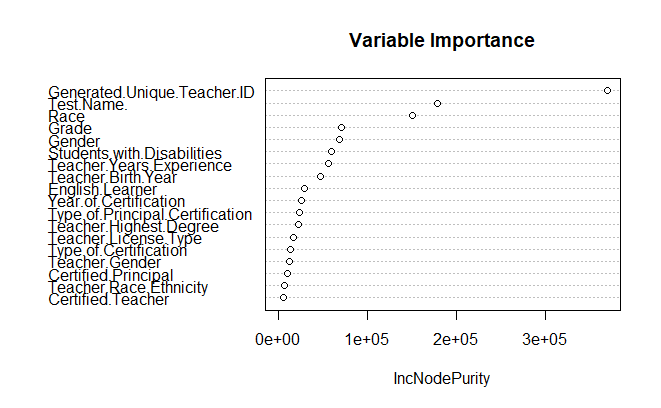
**Data Visualizations**

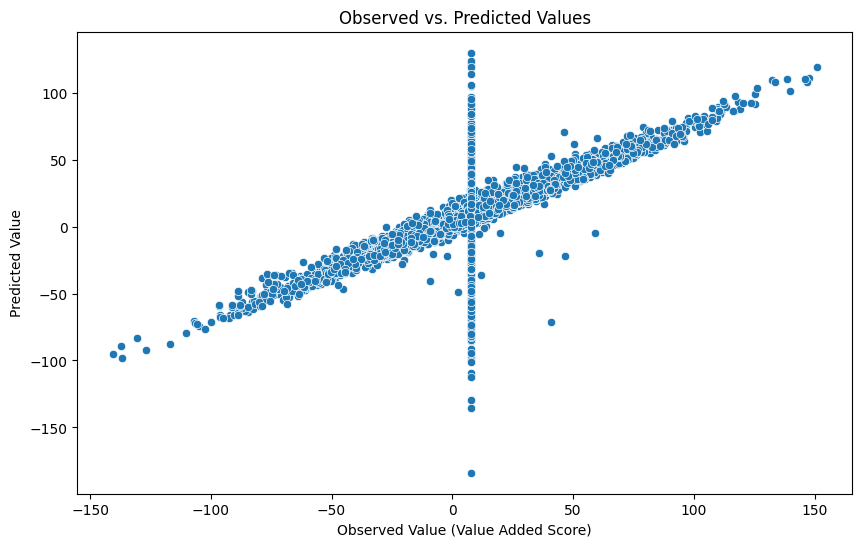
*Include a small sample of relevant data visualizations / summaries for illustration*







*Image X.* An output showing variable importance of a random forest regression predicting Growth, the variable of interest for the math scores.



**Summary of Modeling and Analysis**

With the goal of determining if culturally responsive certifications impact student outcomes, we plan to investigate a variety of models. First, we plan to look at multivariate linear regression, specifically looking at which variables are statistically significant. Additionally, we plan to investigate random forests and decision trees, seeking to understand which variables have a high “variable importance” score. Finally, we plan to conduct statistical significance tests on a variety of variable groupings.

**Future Work Plan**

| **Plan** | **Due** |
| --- | --- |
| Continue to run regressions and try to identify and justify which variables are of high importance. Try to come to a consensus as to what is believed to be directionally influential to an increase in growth. | March 18 |
| Continue to build and evaluate models | March 31 |
| Write report and finalize presentation | April 28-May 3 |

**Potential Concerns [C] and Blockers [B]**

| **Identifier** | **Description** |
| --- | --- |
| C | It may be challenging to achieve any statistically significant results due to having a relatively small dataset and with how unique some of the data points are when subsetting with different features of the dataset. We are also limited in the types of analyses we can perform due to the small dataset.  Our sponsor will be providing an expanded math dataset which will include 4th and 5th grade, which will slightly alleviate this problem, but not solve it. |