Does Student Race Impact Instructional Spending in K-12 Education in Texas Public School Districts?

Adam Billen, Meklit Gebru, and Mary Gens

January 15, 2021

Abstract

A challenge in the United States is racial disparities in education outcomes, which may be affected by school district spending. However, there is a lack of literature that narrowly focus on racial disparities in instructional spending per pupil. This study aimed to answer if student race impacts instructional spending in K-12 education in Texas public school districts. The results indicate that there is no statistically significant relationship between student race and instructional district spending. However, due to significant limitations in this study's design, results should be interpreted with caution and more research should be conducted.

Background

A persistent challenge in the United States is racial disparities in education outcomes (Rothbart, 2020). This is problematic because education outcomes are correlated with many important variables, including future earnings (Card, 1999) and health care status (Hahn and Truman, 2015). One proposed reason for this disparity is the racial disparity in K-12 public school education funding. Over 90 percent of public schools' funding is through local and state governments (U.S. Department of Education, 2017), which can lead to differences in funding levels across states and local school districts. This relates to race in the U.S. because Black and Hispanic residents tend to live in poorer neighborhoods than white residents (Reardon et al., 2015). In terms of school funding disparities, Morgan and Amerikaner (2018) found that school districts with the highest percentage of students of color received 13 percent less local and state funding than districts with the lowest percent. Education funding reforms in the U.S. have attempted to address these funding disparities, though recent research has found relatively small effects (Rothbart, 2020).

Although there is much research on racial disparities in public school funding, little research has been done on racial disparities in school district spending. School spending is also an important variable to study, as school spending may also impact education outcomes (Baron, 2019; Hyman, 2017). Research on this topic generally shows that racial disparities exist in public school district spending (Sosina and Weathers, 2019; Condron and Roseigno, 2008). However, these studies often use a general spending variable, which may include other types of public school spending that may not directly impact student outcomes, such as operational expenditures. Breaking down this variable by types of spending may help researchers better understand where there are spending disparities and how meaningful those differences may be in terms of effects on students. Specifically, instructional spending is an important variable to consider, as it may impact students' outcomes directly than other types of school spending. This is because teachers directly interact with students, so their training and resources may more significantly affect students' outcomes, such as test scores, than other school staff. Analyzing this specific variable is an essential contribution to the literature. As more policymakers understand racial disparities in public school funding and spending, it may impact education outcomes because there is a greater chance that targeted and tailored policy interventions to address this issue will be successful. For these reasons, the question this paper aims to answer is if student race impacts K-12 school districts' instructional spending. 1. Based on the above literature, the research team hypothesized that: 2. An increase in the percentage of Black students in a district will decrease instructional spending per pupil. 3. An increase in the percentage of Hispanic students in a district will decrease instructional spending per pupil. 4. An increase in the percentage of white students in a district will increase instructional spending per pupil.

Data Collection

The research team did not use a national database on school districts' funding to analyze because it would be challenging to infer relationships due to the differences in many variables across U.S. states. This study focused its analysis on one state, Texas, which is a prime state to analyze, as it is not only large (over 1,200 school districts (TEA, 2014)) but also relatively racially diverse (41% white, 40% Hispanic, and 13% Black (U.S. Census, 2019)). Unlike other states, Texas's public school funding has remained relatively flat since 2010 (Swaby, 2019), meaning that older data may still be meaningful to state policymakers. This study utilized a 2011 performance rating data set from the Texas Education Agency (TEA). The data is annually collected and publicly released by the TEA through school district reporting (TEA, 2020). The research team used a specific dataset and codebook created by Seung-Ho An from the University of Arizona and Nathan Favero from American University. This dataset was used because it was already cleaned and ready for analysis. It was collected from Mary Gens' American University Quantitative Methods Class Blackboard. The team obtained permission from the authors to use and publish the data on GitHub.

Study Design

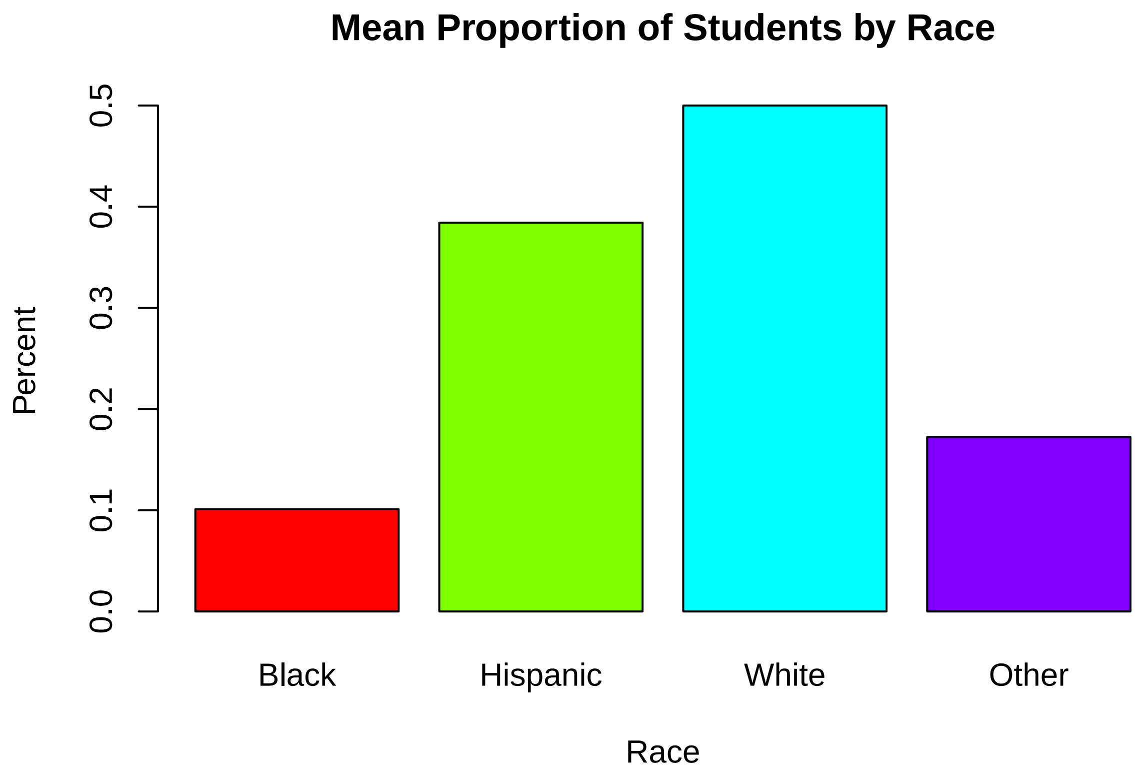
The independent variables for this study were percent of Black students in a district, percent of Hispanic students in a district, and percent of white students in a district. The dependent variable was instructional spending per student. As these variables are all continuous variables, a linear regression was implemented with controls. The linear regression model was used for this research to analyze the relationship between the percent of students' race and instructional spending per student in Texas school districts. To minimize the effect of confounding variables, the model controlled the following variables: number of enrolled students in a district, percent of low-income students in a district, and student-teacher ratio in a district. The number of enrolled students was used to control for a size of a district, while percent of low-income students was used to control for student family income. Student-teacher ratio was used to control for resources available to a school. Data was analyzed using R.

Regression Equation: lm(formula = instpup ~ pbstud + phstud + pwstud + enroll + stear + lowinc, data = mydata) mod <- lm(instpup ~ pbstud + phstud + pwstud + enroll + stear + lowinc, data = mydata) summary(mod)

Results

Bar Plot

This simple bar plot represents the racial makeup of Texas school districts in the data. These bars do not add up perfectly to 100% because they are produced by taking the average percent breakdown of every Texas school district. Notably, the racial category "other" is listed, but the TEA does not specify that category's exact makeup in the dataset. This variable will not be included in the regression model to avoid making conclusions or inferences based on an undefined variable.



Scatter Plots

In these scatter plots the x-variables represent the percent of students of a given race and the y-variable represents instructional expenditures per pupil in U.S. dollars. An abline and a smoothed line with geom(smooth) was placed on top of each scatterplot to show the general relationship between the two variables.

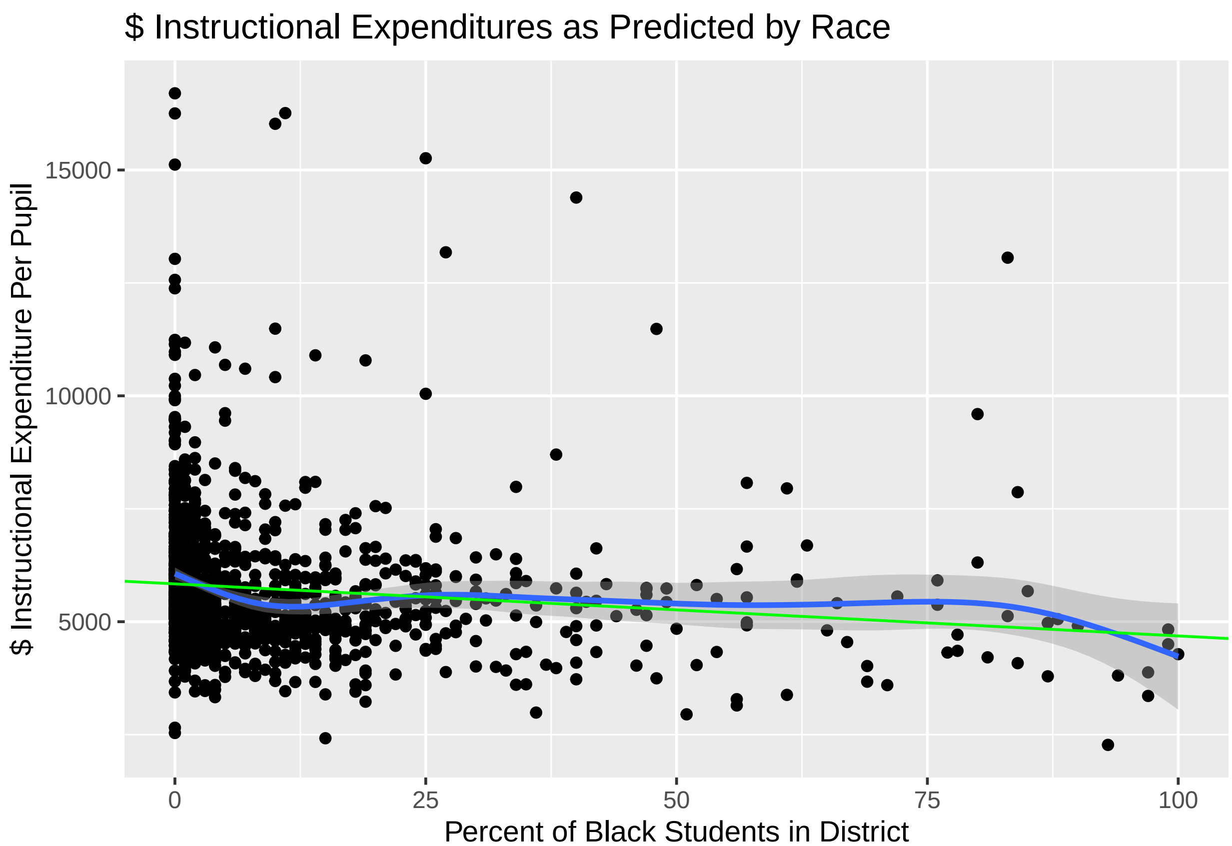
The scatterplot for percent Black students is notable as there is a large cluster of data at x = 0. This means that there is a large number of Texas districts that do not have any Black students. While it was expected that there would be some districts with no Black students due to the racial makeup of Texas in general, the large number above was surprising to the research team. This large number may be the result of misreported data where percent of Black students was unknown and zero was put in its place. This potential limitation must be considered while analyzing the results of this study. The second scatter plot zooms in on this cluster so that it is more easily readable. The abline and geom_smooth line for percent of Black students do show a slight downward trend if the data is taken at face value.

The percent of Hispanic students scatterplot shows a much more predictable pattern. The points are scattered across the plot and show no significant clustering at any particular value and show no clustering at x = 0. This was expected, as Hispanic residents make up 40 percent of the Texas population. The abline and geom_smooth line for Hispanic students are both near perfectly flat and show no visible relationship between racial category and instructional expenditure per pupil.

The percent of white students scatterplot shows a similar pattern to that of the Hispanic student scatterplot. The data points are scattered across the plot and show no intense clustering around a certain value. There are some at points x = 0, but not nearly as many as the scatterplot for percent of Black students. This finding is reasonable as white residents only make up 41 percent of the Texas population. The abline and geom_smooth lines show a very slight upward trend, though this trend is even less significant than that of the downward trend for Black students.

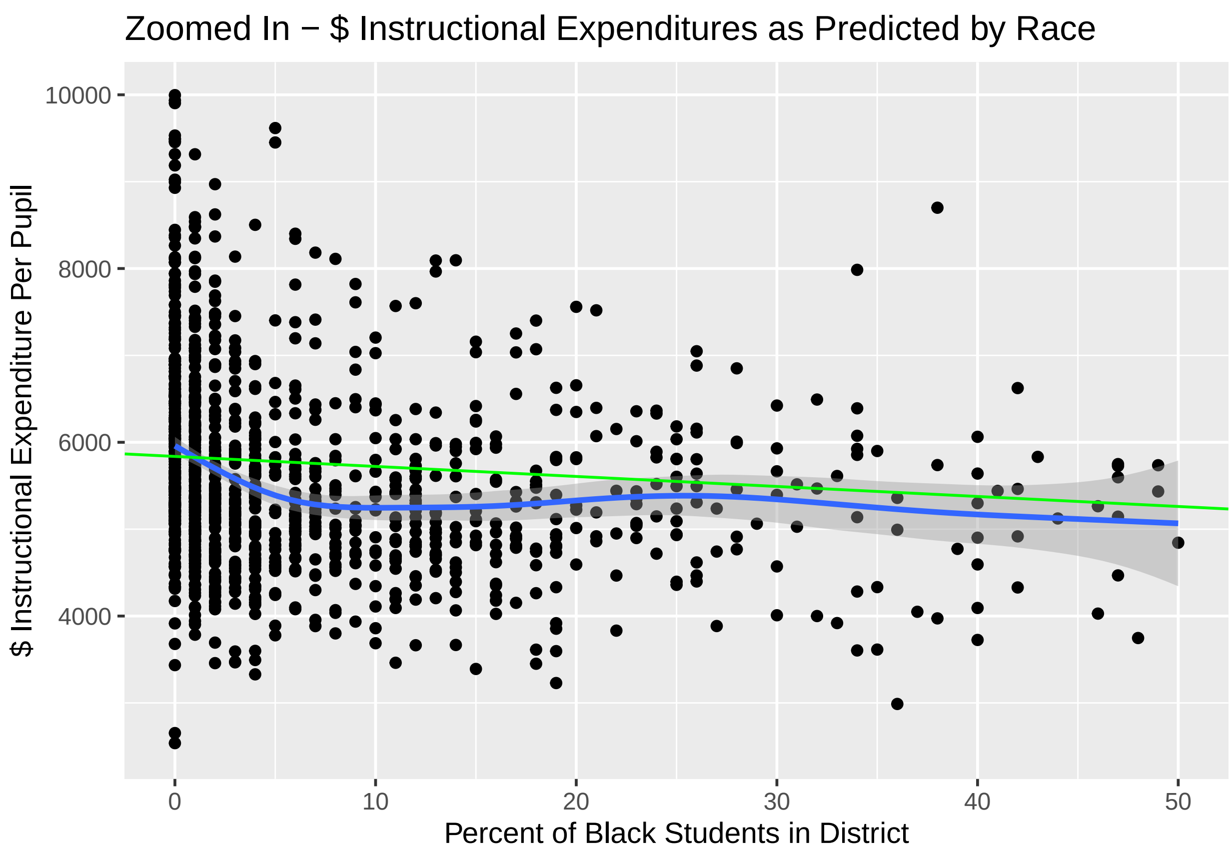
```
## Warning: Removed 89 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 89 rows containing missing values (geom_point).
```



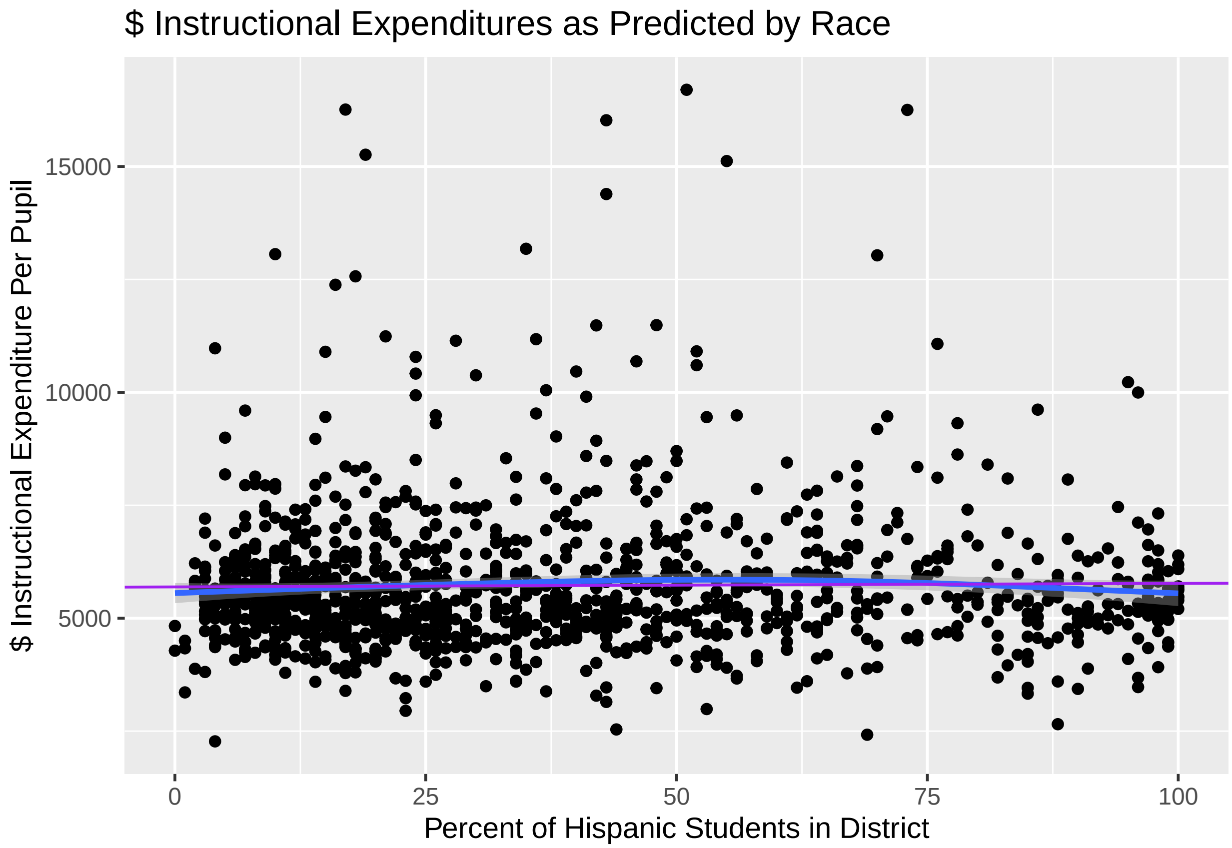

```
## Warning: Removed 167 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 167 rows containing missing values (geom_point).
```



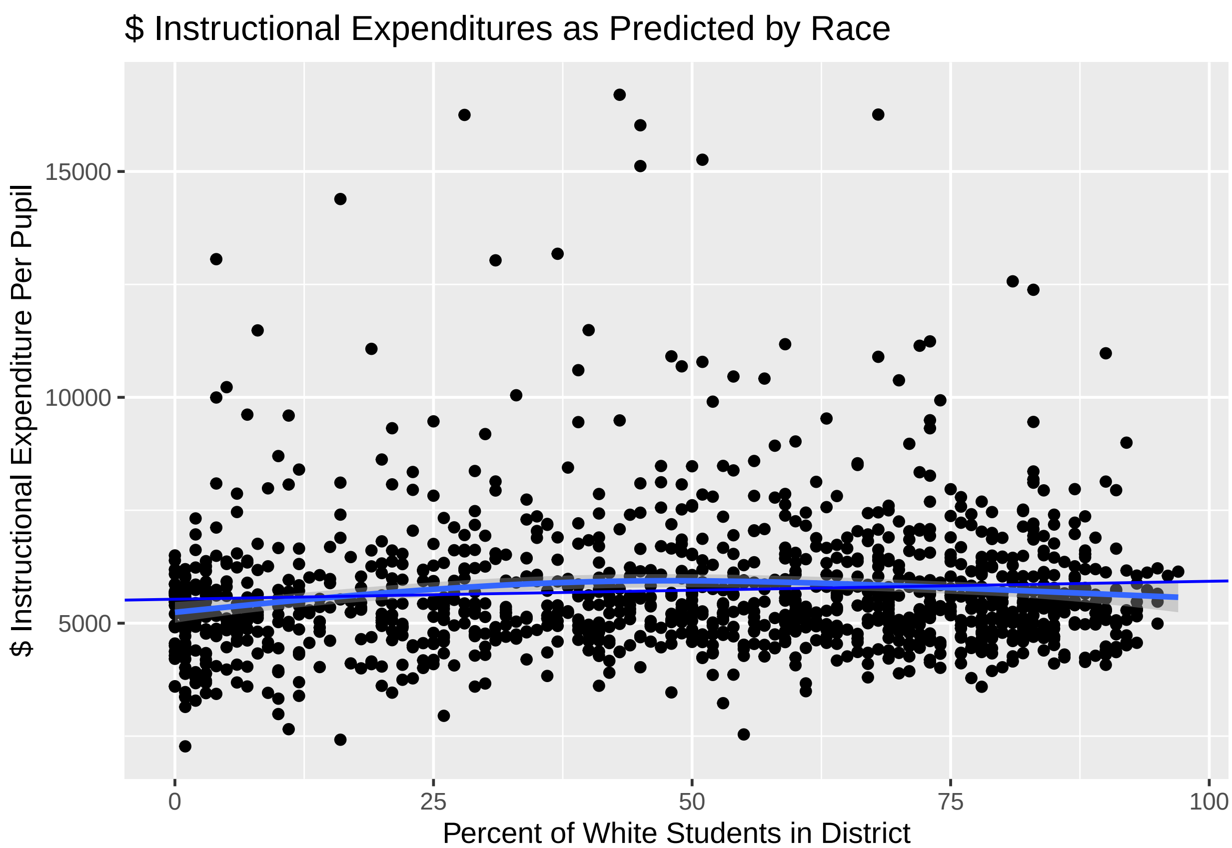

```
## Warning: Removed 89 rows containing non-finite values (stat_smooth).

## Warning: Removed 89 rows containing missing values (geom_point).
```




```
## Warning: Removed 89 rows containing non-finite values (stat_smooth).

## Warning: Removed 89 rows containing missing values (geom_point).
```



Regression Output

Number of observations = 1190 F-statistics= 146.2 on 6 and 1216 DF R-squared = 0.4191 Adj R-squared = 0.4162

The coefficient for percentage of Black students indicates that for every additional percent of Black students in the district, instructional spending per pupil increases by an average of ($4.16) while controlling everything else in the model. However, a high p-value (0.66) indicates that the relationship between percentage of Black students and instructional spending per pupil is not statistically significant. This means there is no association between the percent of black students and instructional spending per pupil. The coefficient for percentage of Hispanic students indicates that for every additional percent of Hispanic students in the district, instructional spending per pupil increases by an average of ($7.24) while controlling for the other variables in the model. The p-value is high (0.407) which indicates that both percentage of Hispanic students and instructional spending per pupil is not statistically significant. This means that there is no association between the percent of hispanic students and instructional spending per pupil. The coefficient for percentage of white students indicates that for every additional percent of white students in the district, instructional spending per pupil decreases by an average approximately ($0.86) while controlling everything else in the model. The p-value is high (0.92) which indicates that both percentage of white students and instructional spending per pupil is not statistically significant. This means that there is no association between the percent of Hispanic students and instructional spending per pupil. It is notable that the student-teacher ratio variable is statistically significant (p=0.00). This is potentially due to the fact that instructional spending may be largely made up of teachers' salaries. Therefore, more teachers would lead to larger instructional spending.

```
##
## Call:
```

```
## lm(formula = instpup ~ pbstud + phstud + pwstud + enroll + stear +
##     lowinc, data = mydata)
##
## Coefficients:
## (Intercept)       pbstud       phstud       pwstud       enroll        stear
##   9.579e+03    4.160e+00    7.242e+00   -8.608e-01    1.832e-03   -3.427e+02
##      lowinc
##   3.180e+00
```

```
##
## Call:
## lm(formula = instpup ~ pbstud + phstud + pwstud + enroll + stear +
##     lowinc, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3781.6  -577.8  -200.6   256.6  8287.0
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.579e+03  8.467e+02  11.314   <2e-16 ***
## pbstud       4.160e+00  9.331e+00   0.446    0.656
## phstud       7.242e+00  8.724e+00   0.830    0.407
## pwstud      -8.608e-01  8.723e+00  -0.099    0.921
## enroll       1.832e-03  2.958e-03   0.619    0.536
## stear       -3.427e+02  1.246e+01 -27.504   <2e-16 ***
## lowinc       3.180e+00  2.427e+00   1.310    0.190
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1190 on 1216 degrees of freedom
##   (90 observations deleted due to missingness)
## Multiple R-squared:  0.4191, Adjusted R-squared:  0.4162
## F-statistic: 146.2 on 6 and 1216 DF,  p-value: < 2.2e-16
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##     0.0     1.0     3.0    10.1    12.0   100.0      85
```

```
##
## Call:
## lm(formula = pbstud ~ instpup, data = mydata)
##
## Coefficients:
## (Intercept)      instpup
##   17.325738    -0.001285
```

```
##
## Call:
## lm(formula = pbstud ~ instpup, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.066  -9.032  -6.304   1.770  88.175
```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 17.3257384  1.7781305   9.744  < 2e-16 ***
## instpup     -0.0012849  0.0002998  -4.285 1.97e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.32 on 1222 degrees of freedom
##   (89 observations deleted due to missingness)
## Multiple R-squared:  0.0148, Adjusted R-squared:  0.014
## F-statistic: 18.36 on 1 and 1222 DF,  p-value: 1.969e-05
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   15.00   31.50   38.42   56.00  100.00      85
```

```
##
## Call:
## lm(formula = phstud ~ instpup, data = mydata)
##
## Coefficients:
## (Intercept)      instpup
##   3.697e+01    2.488e-04
```

```
##
## Call:
## lm(formula = phstud ~ instpup, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -38.169 -23.191  -6.769  17.666  61.738
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.697e+01  3.031e+00  12.197   <2e-16 ***
## instpup     2.488e-04  5.111e-04   0.487    0.626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.83 on 1222 degrees of freedom
##   (89 observations deleted due to missingness)
## Multiple R-squared:  0.000194,   Adjusted R-squared:  -0.0006242
## F-statistic: 0.2371 on 1 and 1222 DF,  p-value: 0.6264
```

```
## Warning: Unknown or uninitialised column: `pWstud`.
```

```
## Length  Class   Mode
##      0   NULL   NULL
```

```
##
## Call:
## lm(formula = pwstud ~ instpup, data = mydata)
```

```
##
## Coefficients:
## (Intercept)      instpup
##   40.581871     0.001355

##
## Call:
## lm(formula = pwstud ~ instpup, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -54.273 -24.332   3.242  25.396  48.099
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.058e+01  3.126e+00   12.98   <2e-16 ***
## instpup     1.355e-03  5.271e-04    2.57   0.0103 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.69 on 1222 degrees of freedom
##   (89 observations deleted due to missingness)
## Multiple R-squared:  0.005378,   Adjusted R-squared:  0.004564
## F-statistic: 6.607 on 1 and 1222 DF,  p-value: 0.01028

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    2273    4860    5405    5722    6182   16698      89
```

Limitations

The dataset utilized in this study is administration data; therefore, it is difficult to create adequate controls for the regression. This is because the initial reason for collecting this dataset was for accountability purposes, so the metrics this study was interested in exploring were not considered during the data collection process. For example, other possible factors that may impact the dependent variable were not included in TEA's data collection, such as parent income or education level. As such, only a limited number of variables could be controlled for. Moreover, the research team could not explore the dataset in-depth due to missing data from how the dataset was collected. For example, the dataset had a lot of missing data which could have potentially impacted our coefficients. For this reason the statistical power for this study was reduced due to missing data and thus the research team might have overestimated the impact of racial disparities on instructional spending due to omitted variables. Finally, as the number of districts that do not have any Black students may not accurately reflect reality, the results of the regression model should be analyzed with caution. In order to increase confidence in the results, future researchers should research whether this data reflects reality and, if not, why.

Analysis

This study's results differ from previous literature in finding that there is no statistically significant relationship between percent of students' given race and instructional spending per pupil. This means that, according to this study, the percent of a students' given race in a district does not lead to a higher or lower instructional spending per pupil in that Texas district. However, limitations of this study may have significantly impacted these results. Therefore, the results of this study should be taken with caution. It is also notable that this study design does not have external validity. These results only aim to explain the relationship between variables in the state of Texas. They should not be generalized to the relationship between analyzed variables in the U.S. as a whole or in other U.S. states. This is because there are many differences between Texas and other U.S. states, and so potentially critical variables were likely not controlled for in this study design.

Conclusion:

Based on the results of this study, the research team fails to reject the null hypotheses. However, due to the limitations of this study and the importance of the topic, more research should be done with a different dataset and/or a different study design. A randomized control trial, in particular, may be a useful model to better understand the relationship between instructional district spending and student race. At this point, policymakers should not use this study's results to inform any education policy reforms.

Works Cited

1. Baron, E. (2019). School Spending and Student Outcomes: Evidence from Revenue Limit Elections in Wisconsin. SSRN. doi:https://dx.doi.org/10.2139/ssrn.3430766
2. Card, D. (1999). The Causal Effect of Education on Earnings. Handbook of Labor Economics, 3, 1801-1863. doi:https://doi.org/10.1016/S1573-4463(99)03011-4
3. Condron, D. J., & Roseigno, V. J. (2003). Disparities within: Unequal Spending and Achievement in an Urban School District. Sociology of Education, 76(1), 18-36. doi:http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.397.4522&rep=rep1&type=pdf
4. Federal Role in Education. (2017, May 25). Retrieved January 14, 2021, from https://www2.ed.gov/about/overview/fed/role.html
5. Hahn, R. A., & Truman, B. I. (2015). Education Improves Public Health and Promotes Health Equity. International Journal of Health Services, 45(4), 657-678. doi:https://dx.doi.org/10.1177%2F0020731415585986
6. Hyman, J. (2017). Does Money Matter in the Long Run? Effects of School Spending on Educational Attainment. American Economic Journal: Economic Policy, 9(4), 256-280. doi:10.1257/pol.20150249
7. Morgan, I., & Amerikaner, A. (2020, February 20). Funding Gaps 2018. Retrieved January 14, 2021, from https://edtrust.org/resource/funding-gaps-2018/
8. Reardon, S. F., Fox, L., & Townsend, J. (2015). Neighborhood Income Composition by Household Race and Income, 1990–2009. The ANNALS of the American Academy of Political and Social Science, 660(1), 78-97. doi:https://doi.org/10.1177%2F0002716215576104
9. Rothbart, M. W. (2020). Does School Finance Reform Reduce the Race Gap in School Funding? Education Finance and Policy, 15(4), 675-707. doi:https://doi.org/10.1162/edfp_a_00282
10. Sosina, V. E., & Weathers, E. S. (2019). Pathways to Inequality: Between-District Segregation and Racial Disparities in School District Expenditures. AERA Open, 5(3). doi:https://journals.sagepub.com/doi/pdf/10.1177/2332858419872445
11. Swaby, A. (2019, February 15). Texas' school finance system is unpopular and complex. Here's how it works. Retrieved January 14, 2021, from https://www.texastribune.org/2019/02/15/texas-school-funding-how-it-works/
12. Texas Education Agency. (2014, August 14). Overview of Texas Schools. Retrieved January 14, 2021, from https://tea.texas.gov/texas-schools/general-information/overview-of-texas-schools
13. Texas Education Agency. (2020). Performance-Based Monitoring. Retrieved January 14, 2021, from https://tea.texas.gov/student-assessment/monitoring-and-interventions/data-validation-monitoring-0
14. U.S. Census Bureau QuickFacts: Texas. (2019). Retrieved January 14, 2021, from https://www.census.gov/quickfacts/TX