An Exploration of the Texas Education Agency 2022-2023 Accountability Ratings Dataset

Radiah Khan, Mercer Mercer, Nafisa Mohamed, Catherine Weeks

Introduction

Through this paper, we aim to address the question: Do schools in Texas with higher percentages of economically disadvantaged students have lower accountability ratings? To address this, we sourced our data from Texas's annual academic accountability ratings, which it provides for all of its public and charter schools. This question is of particular interest now because the Academic accountability ratings are the benchmark the Texas Education Agency (TEA) uses to take over underperforming districts(Texas Education Agency 2025). This adds additional questions to what makes a school at risk for being taken over. The previous two districts to be taken over, Houston ISD and Fort Worth ISD, both had a large percentage of economically disadvantaged students. Some districts have begun redrawing their boundaries to avoid government takeovers(Edison 2025). If schools that predominantly serve economically disadvantaged students are closing to avoid a government takeover, this would cause an undue burden on those students, who would then have to travel further to get to classes, which can already be a barrier to education. Texas has also recently updated the criteria in its accountability rating system, which some teachers and superintendents claim has made earning a passing grade more difficult, potentially affecting whether student economic status influences academic ratings. Previous research has suggested a connection between a student's economic status and school performance. For example, a study from Stanford examining all public school districts in the United States found a strong relationship between district socioeconomic status and average academic achievement (Sean F. Reardon 2016). Another study found a strong connection between student socioeconomic status and student achievement(Jennifer Elizabeth Barry 2006). But neither of these are recent or uses the data that Texas uses to internally select what it considers to be failing schools. Our paper aims to provide additional insight into how Texas's academic achievement ratings relate to students' socioeconomic status, given the state's new academic achievement rating system. Texas's academic accountability rating system reports multiple variables containing information on school performance, but the ones of interest to us are as follows: Academic Growth Score, Overall Score, School Progress Score, and multiple variables indicating distinction in particular areas, which we have combined into a single variable that tracks total earned distinctions.

```
# A tibble: 1,198 x 2
   district                      avg_disadvantaged
   <chr>                                     <dbl>
 1 ACADEMY FOR ACADEMIC EXCELLENCE             100
 2 DRAW ACADEMY                                100
 3 GEORGE GERVIN ACADEMY                       100
 4 LA FE PREPARATORY SCHOOL                    100
 5 STEP CHARTER SCHOOL                         100
 6 PROGRESO ISD                               99.9
 7 SANTA MARIA ISD                            99.9
 8 EXCEL ACADEMY                              99.8
```
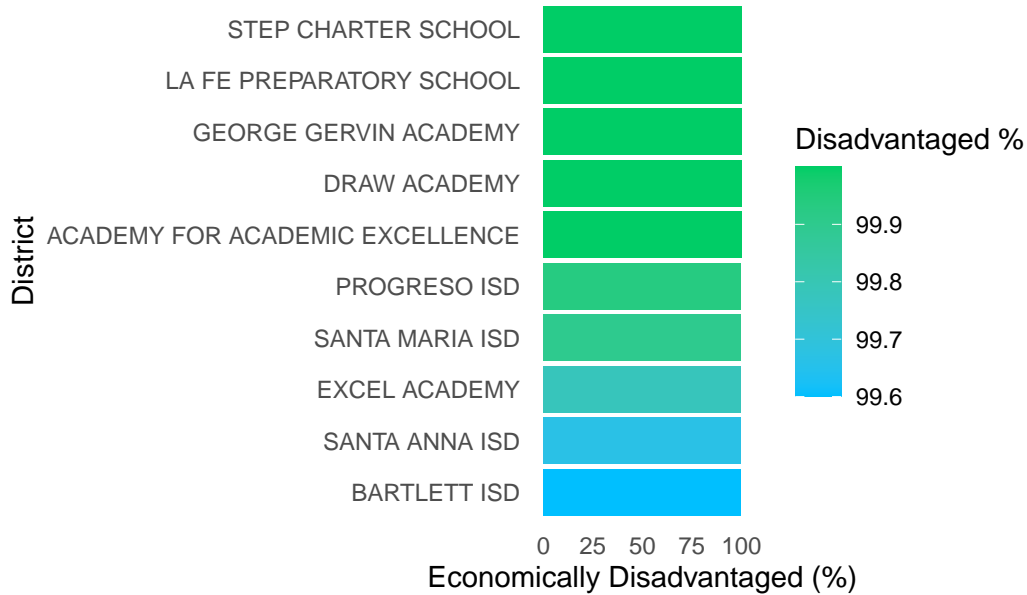
```
 9 SANTA ANNA ISD                                    99.7
10 BARTLETT ISD                                      99.6
# i 1,188 more rows
```

Plot 1

**; with the Highest Average Economically Disadvantaged Student**

The horizontal bar chart of the ten districts with the highest average percent economically disadvantaged highlights a small group of districts and charter schools with very high concentrations of economically disadvantaged students.This shows Schools with 100% or near 100% are a mix between charter and public alternative schools. Additionally, 9 out of 10 of these schools are also within highly metropolitan areas.

Plot 2 "Overall Rating"

Count of Districts by Rating Categories
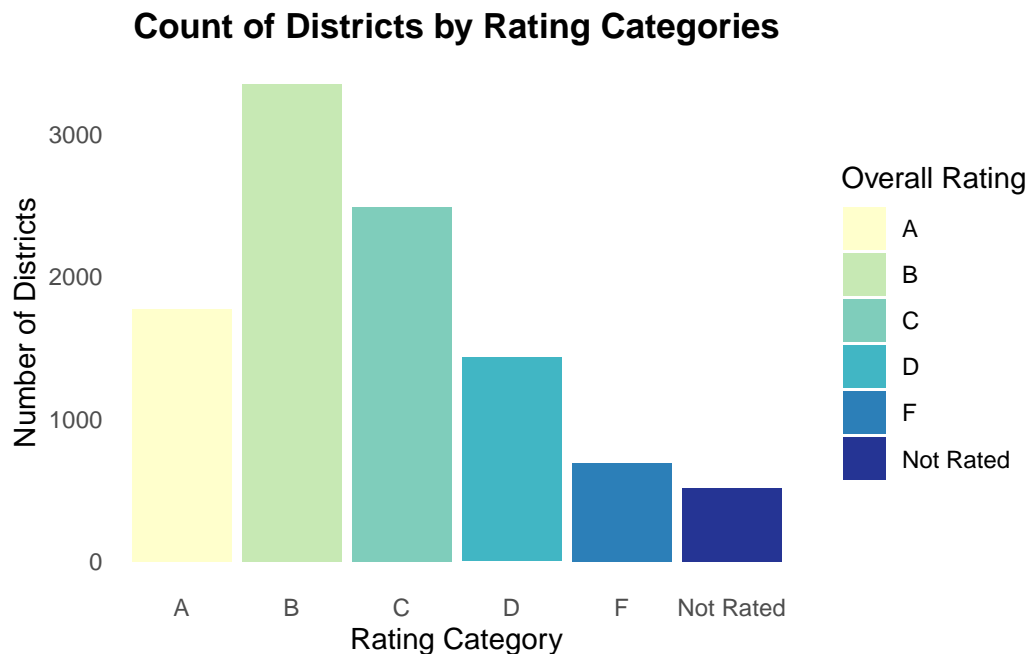
The rating-distribution bar chart shows most districts receive mid-to-high ratings: B and C are the most common categories (with B the largest, followed by C and A), while D, F, and Not Rated are substantially less frequent. In other words, most districts fall into the mid-to-high rating range, creating a central peak with short tails at both the high and low ends. This shape means the distribution is clustered rather than widely spread, so many districts have similar overall ratings and only a minority sit at the extremes. Practically, that implies comparisons based on letter grades will group many districts together and finer distinctions will rely on the underlying numeric scores or other measures to separate performance within that large central mass.

Plot 3 `Academic Growth Score`

Academic Growth and Overall Rating
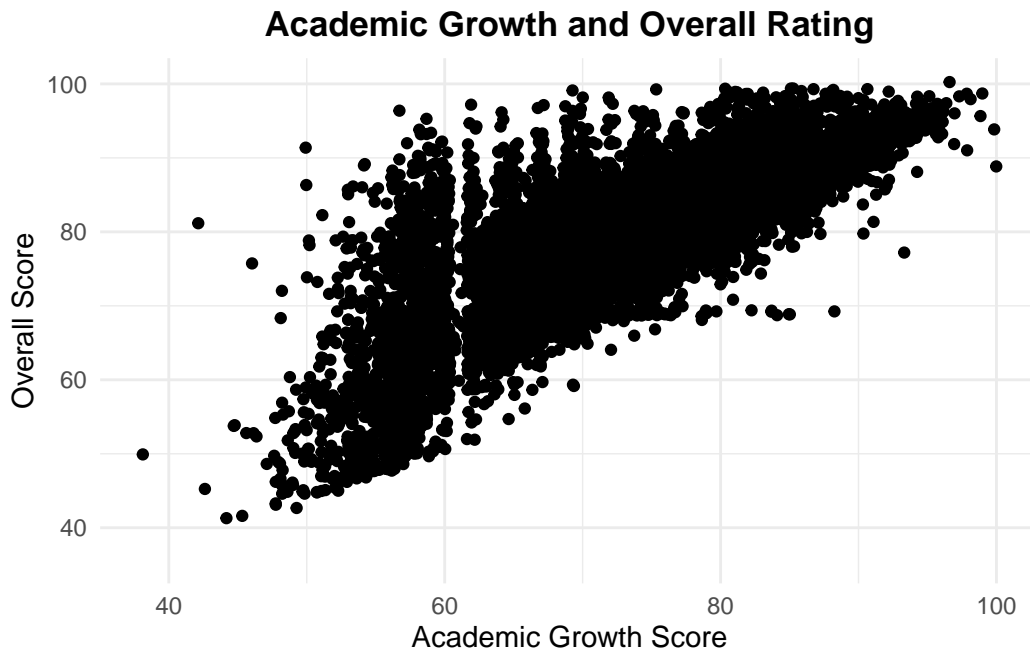
The scatterplot of Academic Growth Score versus Overall Score reveals a clear positive relationship: higher academic growth scores tend to accompany higher overall scores, with many schools clustered along an upward trend from roughly 50–60 overall at lower growth values to 85–100 overall at higher growth values. The spread is wider at lower growth levels and tighter near the top, suggesting a ceiling effect where very high growth scores correspond to consistently high overall scores; this reinforces that academic growth is an important predictor of overall performance, although variability at lower growth levels indicates other factors also influence overall score.

Methods

Data Source

The data set used in this report is from the Texas Education Agency's 2025 State, Region, District, and Campus-Level Accountability Report. The data in this report originates from public and charter schools in Texas overseen by the Texas Education Agency (TEA), who are required to submit all student standardized tests and other academic performance data (Texas Education Agency 2025). Each observation in this data set represents a different campus.

Before building our model, we performed some data pre-processing steps to clean and organize the initial data set for analysis:

Dataset Processing Steps

1. The dataframe had column names that were either capitalized inconsistently/had multiple spaces in between words; which made parsing the names harder for R to work on. So we cleaned the column names. (e.g., `School Progress Score` → `school_progress_score`)

2. The data was filtered to grades 09-12 using `grades_served` column to keep our results consistent and specific to only high school students.

3. There were 7 columns of distinction for each school (Distinction ELA/Reading, Distinction Mathematics, Distinction Science, Distinction Soc Studies, Distinction Progress, DIstinction Closing the Gaps, Distinction PostSecondary Readiness). All of them had values of `Earned`, `Not Earned`, `NA`. We reshaped this wide data into long data and merged it into one single column named `distinction_status`.

4. However, after creating the distinction_status column, each school had seven rows. To prevent this from creating inconsistencies in the results of the chosen model later, we created an additional column, `distinction_score`, which counts the number of times each school earned distinction out of seven and assigns a single score to each school.

5. We removed all the rows with NA values.

6. While working on this data set, the numeric variables had their values as characters and not numeric. So we converted them to numeric values for building our model.

The final variables selected for analysis are:

- `number_of_students`
- `academic_growth_score`
- `percent_economically_disadvantaged`
- `overall_score`
- `school_progress_score`

- `distinction_score`
- `campus`
- `district`

Data Observations Flow

Initial dataset: 10,253 observations

↓

Filtered to grades 9–12: 1,394 observations

↓

After NA removal: 1,356 observations

Variable Units and Range

The original data contains many variables related to school performance, but in this paper we will focus on schools' Overall Score, their Academic Growth Scores, their Distinction Scores, the number of students, and the percentage of students who are economically disadvantaged. Before selecting our model, we chose one variable from here as response variable and the rest as explanatory variable.

Table 1: Response Variable Information

Variable	Type	Range
Overall Score	Numeric	0–100 points

Table 2: Explanatory Variable Information

Variable	Type	Unit of Observation	Range
Number of students	Numeric	Students	1–5,317
Academic growth score	Numeric	School	0–100
Percent economically disadvantaged	Numeric	School	0–100%
School progress score	Numeric	School	0–100
Distinction score	Numeric	School	0–7

Variable Descriptions

1. `Overall Score` : This variable represents the school's final accountability rating on a 0-100 scale. It represents a schoolwide summary of academic performance, progress, and equity. This score is calculated by using three accountability domains: Student Achievement, School Progress, and Closing the Gaps. For this score, 70% is made up of either the Student Achievement or School progress (it picks the "better outcome") and 30% of it is from their Closing the Gaps domain score (which evaluates how well a school is supporting the performance of specific students, specifically economically disadvantaged students, students from certain racial groups, and special needs students among others). The data handbook further explains how these were specifically calculated (Texas Education Agency 2025).

2. `Academic Growth Score` : This variable is a measure of how much students improved academically over time. It is calculated using a raw growth score which is based on whether students met, exceeded, or fell short of STAAR progress (which is further defined in the data handbook). This raw growth score is then converted to a scaled score using a table, which then becomes the academic growth score, which is also measured on a 0-100 scale.

3. `Distinction Score` : This variable is derived from the 7 categories of Distinction that schools report. We explained this derivation in our dataset processing steps. But first here is what distinction means across the categories: It is calculated by comparing a school's performance on multiple indicators to a group of similar campuses. If it places in the top quartile on at least 50% of indicators (for elementary and middle school) or 33% (for high school), it received the distinction. This means that the distinction score is whether or not the school performed exceptionally well compared to similar schools, and is a binary value of 0 (no distinction) or 1 (distinction) across multiple categories. So the distinction score variable sums up the amount of category a school got distinction and reports it on a 0-7 scale.

4. `Percent economically disadvantaged` : The percentage of students who are economically disadvantaged represents the percentage of students who are economically disadvantaged according to a few criteria. These include eligibility for free and reduced lunch prices and SNAP/TANF participation.

5. `Number of students` : This variable represents the number of students enrolled in each campus.

6. `School Progress Score` :

Model Specification

We built a multiple linear regression model using nested F test model.

The reduced model:

$$\widehat{\text{overall score}}_i = \beta_0 + \beta_1 \cdot \text{academic growth score}_i + \beta_2 \cdot \text{distinction score}_i + \beta_3 \cdot \text{school progress score}_i$$

term	estimate	std.error	statistic	p.value
(Intercept)	-3.6260319	1.0890561	-3.329518	0.0008953
academic_growth_score	0.0279850	0.0138865	2.015269	0.0440900
distinction_score	0.3679789	0.0566432	6.496430	0.0000000
school_progress_score	1.0093466	0.0176877	57.064921	0.0000000

The full model:

$$\widehat{\text{overall score}}_i = \beta_0 + \beta_1 \cdot \text{academic growth score}_i + \beta_2 \cdot \text{distinction score}_i + \beta_3 \cdot \text{school progress score}_i$$
$$+ \beta_4 \cdot \text{percent economically disadvantaged}_i + \beta_5 \cdot \text{number of students}_i$$

term	estimate	std.error	statistic	p.value
(Intercept)	7.1637191	1.1783168	6.079621	0.0000000
academic_growth_score	-0.0544489	0.0125998	-4.321398	0.0000167
distinction_score	0.3822513	0.0502416	7.608270	0.0000000
school_progress_score	1.0202478	0.0160914	63.403371	0.0000000
percent_economically_disadvantaged	-0.0889648	0.0042717	-20.826318	0.0000000
number_of_students	-0.0001981	0.0000936	-2.115238	0.0346073

Evaluating assumptions of the model

The four assumptions we need to evaluate are : Linearity, Independence, Normality, Equal Variance.

Assuming independence was maintained, we check the other three conditions. However, we discussed more about the concerns of independence in the discussion section (Section)

For checking the linearity and equal variance we check the residual plot of the reduced and full model. See the detailed plot in Appendix A.

For both of the reduced and full model, it violates the linearity, equal variance and normality condition. From the residual plot, we see that it widens as the x gets bigger and there is a pattern above and below y= 0 line. There seems to be unusual points as well which we will investigate later.

In the Q-Q plot, for the both of the model, the points deviates from the reference line. So it does violate the normality condition.

Model Selection

To determine the model, we started with calculating the VIF to detect any multicollinearity issues among the explanatory variables. We calculated this using both the reduced and full model. There were no multicollinearity issues given all of the VIFs show little collinearity < 5 (see Appendix B for detailed results).

We began to test our model using forward model selection with AIC. See Appendix C for detailed results. According to the results,the AIC drops when we add `number of students`, it further drops when we add `percent economically disadvantaged` as well.

To find the influential points, we first filtered the data to find the schools with leverage that was outside of the acceptable range, then further filtered that list to the schools that additionally had a standardized residual that was outside of the range. After filtering, we were left with only five schools that were potential influential points. We refit the model without them and saw no noticeable difference, so we added them back into the final data set. See Appendix D for detailed results.

Results

Interpretation of the fitted model

Interpretation of the relevant co-effecients

Practical Importance

Corresponding Confidence Interval

```
vcov(full_model)
```

```
                                   (Intercept) academic_growth_score
(Intercept)                       1.3884304498         -3.144774e-03
academic_growth_score            -0.0031447742          1.587559e-04
distinction_score                 0.0274063815         -1.079823e-05
school_progress_score            -0.0122210799         -1.176643e-04
percent_economically_disadvantaged -0.0024101108         1.647127e-05
number_of_students               -0.0000508727         -1.486157e-08
                                  distinction_score school_progress_score
(Intercept)                            2.740638e-02         -1.222108e-02
academic_growth_score                 -1.079823e-05         -1.176643e-04
distinction_score                      2.524214e-03         -3.650295e-04
```

```
school_progress_score                  -3.650295e-04        2.589325e-04
percent_economically_disadvantaged     -8.300171e-06        1.073369e-07
number_of_students                     -1.129537e-06        4.768718e-07
                                   percent_economically_disadvantaged
(Intercept)                                          -2.410111e-03
academic_growth_score                                 1.647127e-05
distinction_score                                    -8.300171e-06
school_progress_score                                 1.073369e-07
percent_economically_disadvantaged                    1.824785e-05
number_of_students                                    8.195769e-08
                                   number_of_students
(Intercept)                              -5.087270e-05
academic_growth_score                    -1.486157e-08
distinction_score                        -1.129537e-06
school_progress_score                     4.768718e-07
percent_economically_disadvantaged        8.195769e-08
number_of_students                        8.769878e-09
```

```
g <- coef(full_model)[1] + coef(full_model)[5]
g_var <- vcov(full_model)[1,1] + vcov(full_model)[5,5] + 2*vcov(full_model)[1,5]
g_se <- sqrt(g_var)
t_stat <- (g- 0)/g_se; as.numeric(t_stat)
```

```
[1] 6.014529
```

```
qt(0.975,1255 - 5 -1)
```

```
[1] 1.961865
```

```
upper_ci <- (7.07454 + 1.961865*1.76277)
lower_ci <- (7.07454 - 1.961865*1.76277)
lower_ci
```

```
[1] 3.616223
```

```
upper_ci
```

```
[1] 10.53286
```

Interpret the hypothesis test and its' results - done

The full model contains 5 predictor terms and the reduced model contains 3 predictor terms. \

$$
H\_0:\ \beta\_j = 0 \quad \text{for all p predictor terms being dropped from the full model }
$$
$$
H\_A:\ \beta\_j \neq 0 \quad \text{for atleast one of the p predictor terms being dropped from
$$
We can conduct this nested F tests using anova.

::: {.cell}

```{.r .cell-code}
anova_test <- anova(reduced_model, full_model)
broom::tidy(anova_test) |>
  kableExtra::kable()
```

term	df.residuals	rss	df	sumsq	statistic	p.value
overall_score ~ academic_growth_score + distinction_score + school_progress_score	1251	16542.06	NA	NA	NA	NA
overall_score ~ academic_growth_score + distinction_score + school_progress_score + percent_economically_disadvantaged + number_of_students	1249	12243.08	2	4298.97	219.28	0

:::

We reject the null hypothesis and conclude that we have very strong evidence that atleast one of the extra terms is necessary ($p < 0.05$) taking the significance level 0.05.

Plot

Significance of Predictors
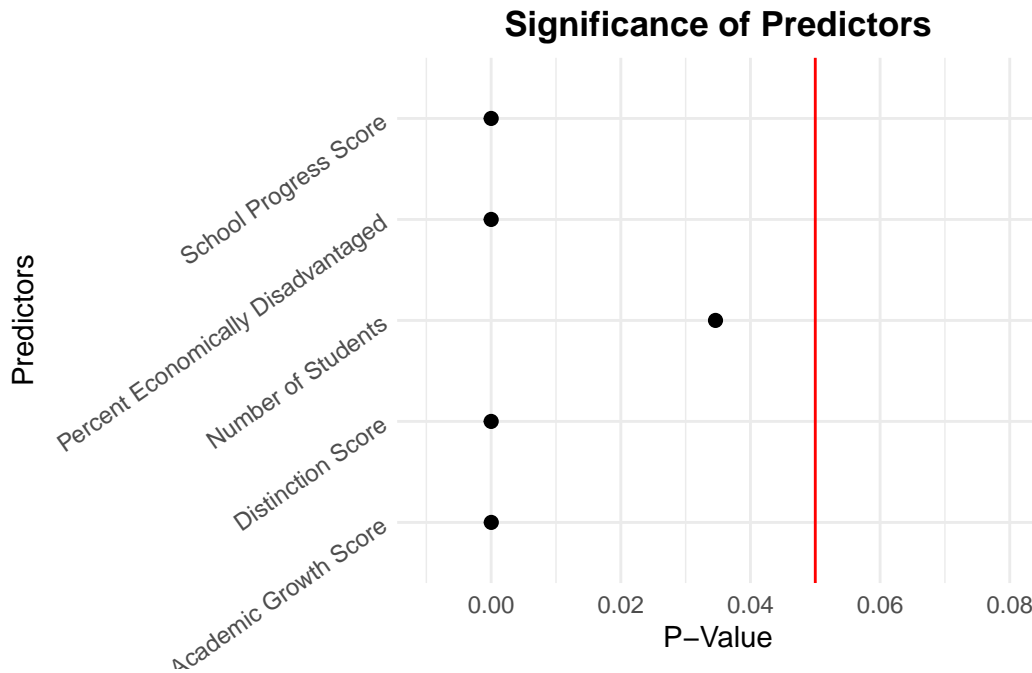
This plot shows us that all of our predictor is lower than 0.05, which is our significance level.

Discussion

The goal of this study was to address the question whether schools in Texas with higher percentages of economically disadvantaged students have lower accountability ratings. Some things should be considered before drawing any conclusions from this model. First, one potential issue with this model is a violation of independence among the variables. Because the Overall Score variable is calculated using some of the metrics within the explanatory variables, we have some concern about this model's independence. However, one of the strengths of this study was a very low VIC and AIC across all variables. So we are somewhat confident that this should not be a major problem, as it indicates the variables are not collinear. Another thing to consider is that this data is only from one year, and it is very difficult to build a dataset with longer term data due to Texas's constant fluctuations in how it reports and calculates the same data across many years. Next, the major strength of this study is that the model performs very well and yields statistically significant results. Finally, the conclusions drawn from this should also only be used to form conclusions about public and charter schools within Texas in the academic year 2024-2025. This is because the metrics used to create many of the variables tested are specific to Texas's academic measurements, such as the STAAR test, and
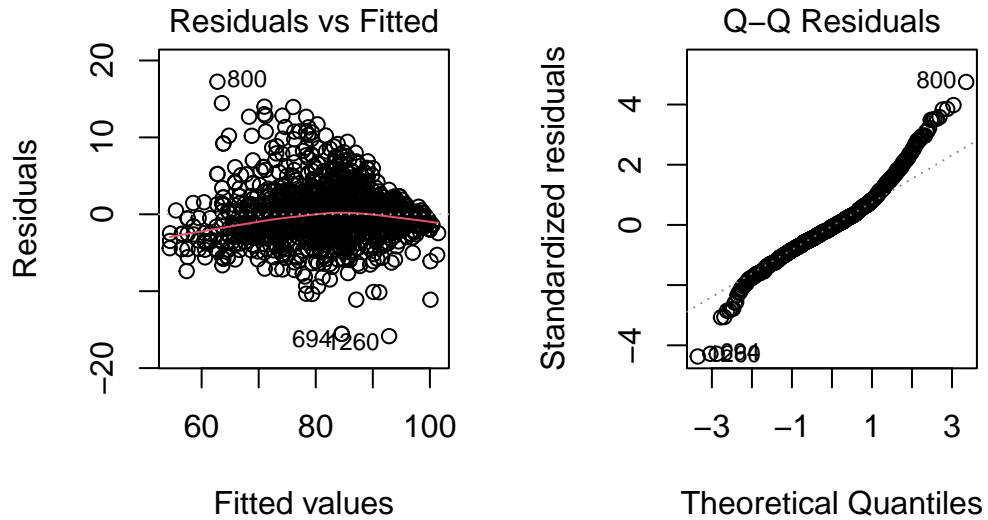
13

cannot easily be compared with those of other States, and the data used is limited to only Texas public and charter schools.
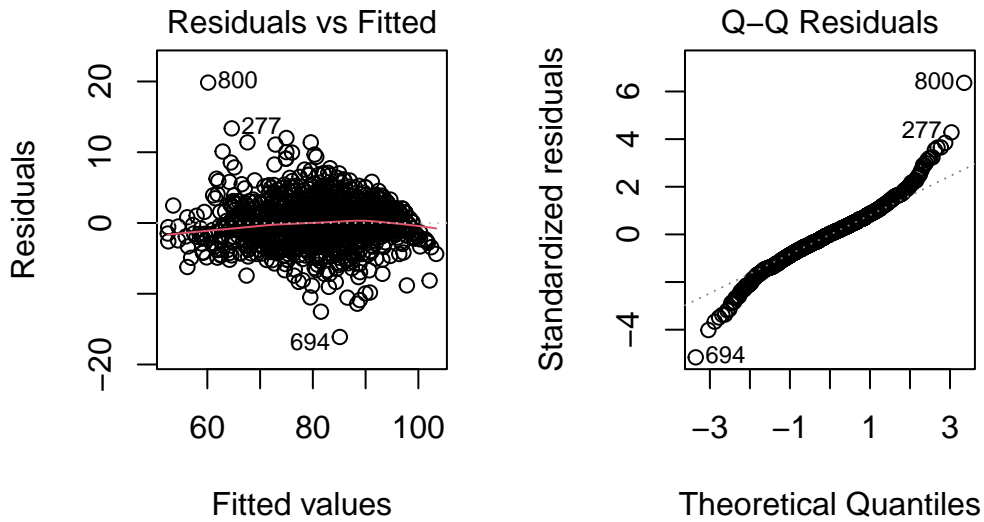
Appendix

Appendix A: Conditions of the linear regression model

We explored the residual and Q-Q plot of both of the model to see if it violates the linearity, nornmality and equal variance condition.

Diagnostic Plots for the Reduced Model

Diagnostic Plots for the Full Model

Residuals vs Fitted

Q–Q Residuals

Appendix B: VIF

Checking the VIF for the full model.

```
        academic_growth_score                           distinction_score
                     2.119657                                    1.498755
        school_progress_score percent_economically_disadvantaged
                     2.554894                                    1.210145
            number_of_students
                     1.197254
```

Appendix C: Stepwise AIC

Checking to see how adding additional parameters affect AIC.

```
Single term additions

Model:
overall_score ~ academic_growth_score + distinction_score + school_progress_score
                              Df Sum of Sq   RSS    AIC  F value  Pr(>F)
<none>                                      16542 3244.4
```

```
number_of_students                     1      47.4 16495 3242.8    3.5893 0.05838
percent_economically_disadvantaged  1    4255.1 12287 2873.2 432.8900 < 2e-16

<none>
number_of_students                  .
percent_economically_disadvantaged ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix D: Influential Points

This is the result of the summary table after we removed the points. As we can see the conclusion does not change.

```
Call:
lm(formula = overall_score ~ academic_growth_score + distinction_score +
    school_progress_score, data = final_df)

Residuals:
    Min      1Q   Median      3Q      Max
-15.8502  -2.0717  -0.2918   1.7927  17.2446

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          -3.66095    1.09291  -3.350 0.000833 ***
academic_growth_score 0.02766    0.01393   1.986 0.047271 *
distinction_score     0.36778    0.05674   6.482  1.3e-10 ***
school_progress_score 1.01009    0.01777  56.852  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.642 on 1247 degrees of freedom
  (135 observations deleted due to missingness)
Multiple R-squared:  0.8723,    Adjusted R-squared:  0.872
F-statistic:  2839 on 3 and 1247 DF,  p-value: < 2.2e-16
```

Edison, Sneha Dey, Jaden. 2025. "Texas School Leaders: State Takeovers Are Not Sustainable." https://www.texastribune.org/2025/11/15/texas-tribune-festival-state-takeovers-vouchers/.

Jennifer Elizabeth Barry. 2006. "The Effect of Socio-Economic Status on Academic Achievement." PhD thesis, Department of Sociology.

Sean F. Reardon. 2016. "School District Socioeconomic Status, Race, and Academic Achievement," April.

Texas Education Agency. 2025. "2025 Accountability Manual," March.