The Relationship of Student Attendance and Student Proficiency in English, Language, and Arts

Tram Anh Hoang, Haitham Anbar, Maria Coronado, Erick Njue

2022-12-05

Abstract

Introduction

We use data from the Oregon Department of Education (ODE) to examine the following research questions:

- Does attendance rate affect student proficiency in English, Language, and Arts? and;
- Does such a relationship between attendance and proficiency vary for different racial and ethnic groups?

We use data of student attendance rate and student proficiency rate from the ODE assessment group reports for all the schools in the state. We will base our assessment on the latest academic year available for both data sets, which is 2021 − 2022.

Methods

In this research, we run a regression model on attendance data and proficiency data of schools in Oregon. We obtain two different data sets from the ODE: attendance rate by school and student group, and proficiency in English, Language, and Arts by student group.

Data description and preparation

We use (R Core Team 2021) package (Müller 2020) to read in the two data sets from Oregon Department of Education (ODE)into the (R Core Team 2021) environment. We use (Firke 2021) to clean the names. After preparing the data sets, we display them as tables using (Wickham et al. 2019).

The first data set is student performance in English Language and Arts (ELA) from all schools in the state. The data set provides the number and percentage of students that are proficient in each of the four levels of ELA proficiency, stratified by school, race and ethnicity, and grade level. This data set has 20046 rows and 20 columns.

The second data set is student attendance data from all the schools in Oregon. This data set provides the number and percentage of regular attenders as well as those of chronically absent students, stratified by school and either race and ethnicity or grade level. This data set has 42295 rows and 11 columns.

We clean the two data sets by removing rows with no data and rows with suppressed data, marked as "- -" and "*" respectively. We use {dplyr} package functions in (Wickham et al. 2019) to do so.

After cleaning, we merge the two data sets by matching schools' IDs and student race and ethnicity. Rows with attendance data only or proficiency data only will be removed. We use the *left_join* in the {dplyr} and *drop_na* in{tidyr}, all contained in (Wickham et al. 2019)

Methods

Research question, hypothesis Details about sample: sample size =n, race and ethnic group

Our paper looks at two research questions:
1. Does attendance rate affect student proficiency in English, Language, and Arts?; and
2. Does such a relationship between attendance and proficiency vary for different racial and ethnic groups?

Our data sets obtained from the Oregon Department of Education is school-level data, broken down by racial and ethnic groups. While the relationship between attendance and student proficiency would be better explained from student-level data, we were not able to obtain data in such format.

RQ1: Does attendance rate affect student proficiency in English, Language, and Arts?

Our hypothesis is that schools with higher student attendance have higher rate of student proficiency in English, Language, and Arts.

To answer the first research question, we run a linear regression model, using schools' proficiency rate as the dependent variable, and attendance rate as the independent variable. Our regression equation is: percent_proficient = a + b*percent_regular_attenders + e

RQ2: Does such a relationship between attendance and proficiency vary for different racial and ethnic groups?

Once our hypothesis for the first research question is confirmed, we would like to look at whether students' racial and ethnic backgrounds had implications on the effect of attendance on student proficiency. In other words, if students have different racial and ethnic backgrounds, would attendance improve proficiency in the same way. Our hypothesis is that schools will see different impacts of attendance on proficiency for different student groups.

Results

We use the function *pivot_wider* by student group in (Wickham et al. 2019) package {tidyr} to have columns by race. This is followed by *pivot_longer* and another *pivot_wider* to make the table presentable. We use the *kable* function found in (Xie 2021) to present the results.

To interpret our results, we further present them in a scatter plots using {ggplot2} of (Wickham et al. 2019). In Plot 1, we plot the percentage of proficient students based on the rate of regular attendance. In Plot 2 below, we use *facet_wrap* function of {ggplot2} plot the rate of regular attendance by percentage of proficient students based on the race.

Summary table

Table 1: Summary table

| Race/ ethnicity	Percent Chronically absent	Percent proficient
Asian	86.21	64.18
Black African American	50.61	19.60
Hispanic/latino	56.12	26.11
Multi Racial	66.25	51.56
Pacific Islander	36.18	15.99
White	67.28	50.98

A simple regression analysis was run to examine the relationship between the percentage of regular attenders and the percentage of proficient students in Oregon schools. A total of 1086 Oregon schools with existing data were included in the sample. The results indicated that the percent of regular attendance could statistically significantly predict the percent of proficient students B1=0.423, SE(B1)=0.013, t(2237)=32.5, p<.0001. The results explained that regular attendance accounted for 32% of the explained variability in the percent of proficient students, F (1, 2237) = 1056, p < .0001. The regression equation was:

Predicted percentage of proficient students = 46.5 + 0.423 (percentage of regular attendance)

Plot 1

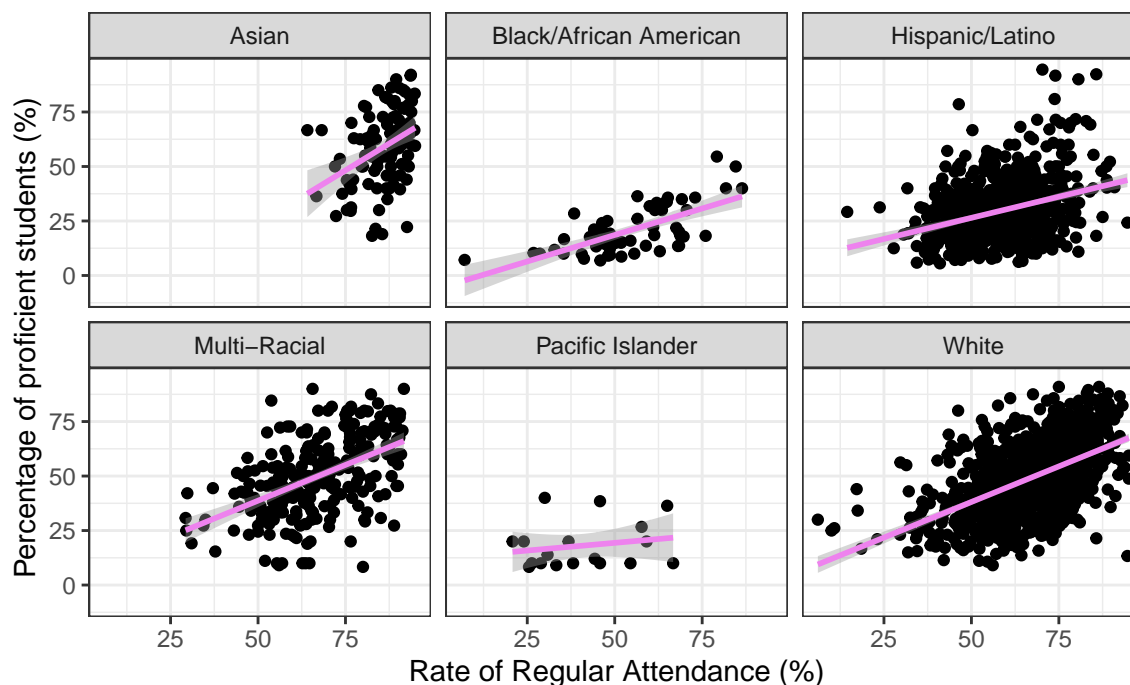Student Attendance and Proficiency Rate
in English, Language, and Arts
Data from Oregon Education Department
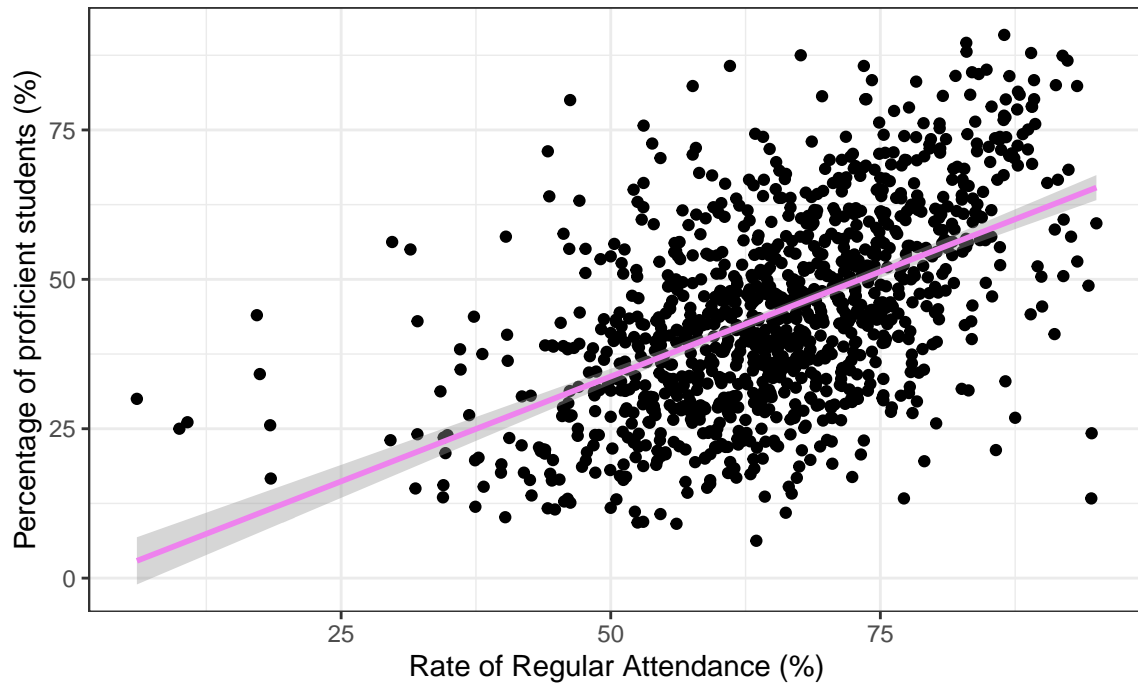
Plot 2

Student Attendance and Proficiency Rate
in English, Language, and Arts by Different Races

Data from Oregon Education Department

Discussions

Regression and Plot 1

When looking at school-level data, attendance rate and proficiency rate has a positive relationship. Schools with 1% higher attendance rate saw a x% increase in their percentage of proficient students. Our linear model explains . . . % of the data point, with the standard error of. . .

Regression and Plot 2

When faceted by students' racial and ethnic groups, we can see a clear change in the slope of the line. Using White students as the reference group, the Asian group has a steeper slope, suggesting 1% increase in attendance rate results in a higher increase in the rate of proficiency for the Asian group than for the White group. Whereas, the Black/African American group, the Hispanic/Latino group, and the Pacific Islander group had flatter line, suggesting that the positive impact of attendance rate on proficiency rate is weaker for these groups than for their White counterparts.

Interestingly, all the data points for the Asian group show up after in the far right part of their plot, indicating high attendance rate for this group of students - above 60% in all the schools we looked at. Contradictingly, the Pacific Islander group had low attendance - below 70% across all schools.

References

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2021. "Knitr: A General-Purpose Package for Dynamic Report Generation in R." https://yihui.org/knitr/.