

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

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

2022-12-06

## Abstract

## Introduction

It is a common assumption that with chronic absenteeism, students' academic performance decreases, as when students fail to attend school, they miss the opportunity to learn. Therefore,. Previous reviews have stated that chronic absenteeism produces multiple impacts on student outcomes, including lower scores on reading and math, social-emotional development, grade retention and dropout and student discipline (Patnode, Gibbons, and Edmunds 2018). On the long run, chronic absenteeism may lead to failure to graduate, underemployment, and financial instability (Henderson, Hill, and Norton 2014). In turn, regular attendance to classes should be a predictor of better performance, and lead to an increased likelihood of graduation.

Multiple social and economic factors can affect attendance to school. Among these are economic hardship, food insecurity, housing instability, health care access, failing home and school infrastructure, transportation, under-resourced schools, school climate and culture, and attendance culture (Henderson, Hill, and Norton 2014). In fact, school attendance can vary by student group. A national report found variations in attendance and chronic absenteeism when comparing gender, race/ethnicity and language status, FRPL (free or reduced-price lunch) eligibility (our proxy for poverty status), and IEP (individualized education program) status (our proxy for disability status) (Garcia and Weiss 2018).

School profiles and reports to the Department of Education include relevant information on school characteristics that can be used to understand how social and economic factors affect student performance (Oregon Department of Education n.d.).In the state of Oregon, previous data has shown how factors related to absenteeism affect student performance. For example, a report from 2014 showed that in Oregon children of color were disproportionately affected by the consequences of chronic absenteeism. In addition, it showed that for 2009-2010, the student groups with higher absenteeism rates were Black, American Indian and Economically Disadvantaged students (Henderson, Hill, and Norton 2014).

Regarding attendance data, in Oregon students are considered “Regular Attenders” if they attended more than 90% of their enrolled days between the beginning of the school year and the first school day in May. In contrast, they are considered chronically absent if they attended 90% or fewer of their enrolled days in this period (Oregon Department of Education 2022).

To explore if regular attendance to classes affected student performance during the school year 2021-2022, we used 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?

# Methods

We ran a regression model on attendance data and proficiency data of schools in Oregon. We obtained two different data sets from the Oregon Department of Education (ODE): attendance rate by school and student group, and proficiency in English, Language, and Arts by student group.

## Data description and preparation

We used (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 cleaned 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 merged 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 used the *left\_join* in the {dplyr} and *drop\_na* in {tidyr}, all contained in (Wickham et al. 2019).A

## Methodological approach

The data sets we obtained from the Oregon Department of Education contain 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. Therefore, all the data presented here is at the school or institution level.

After cleaning the data, a total of 1086 schools were included in the sample, containing information of the percentage of proficient students and the percentage of regular attendance by race or ethnicity. Six races were identified in the data from ODE: Asian, Black African American, Hispanic/latino, Multi Racial, Pacific Islander and White. Pacific Islander included students identified as having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands and not Hispanic. *Table 1* contains a summary of this information. We used 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.

The research questions and methodological approach we used to answer each of them are:

### RQ1: Does attendance rate affect 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:

$$\text{Percent of proficient students} = B_0 + B_1 (\text{Percent of regular attendance}) + e$$

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

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

For the second question, we looked 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. From the literature we reviewed, our hypothesis is that schools will see different impacts of attendance on proficiency for different student groups. To answer this question, we only developed an assessment of the data based on multiple graphics, without fitting a statistical model.

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

## Results

We ran a simple regression analysis 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  $B1=0.75$ ,  $SE(B1)=0.02$ ,  $t(2237)=32.5$ ,  $p<.0001$ . 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:

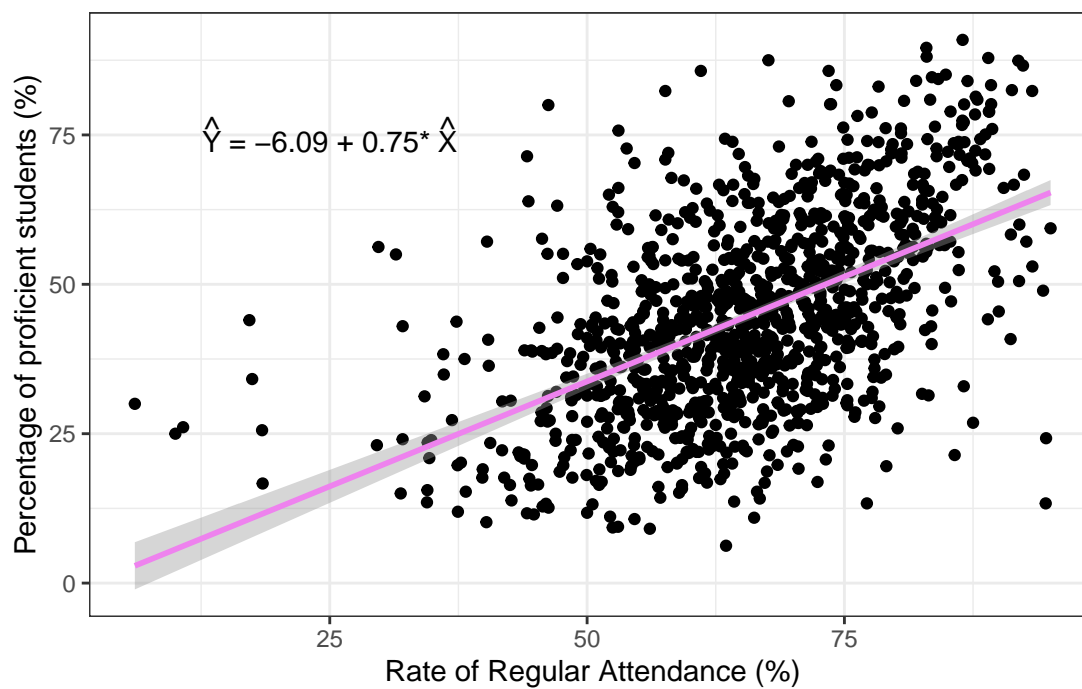
$$\text{Percentage of proficient students} = -6.09 + 0.75 (\text{Percentage of regular attendance})$$

To interpret our results, we created scatter plots using `{ggplot2}` of (Wickham et al. 2019). In Figure 1, we plotted the percentage of proficient students based on the rate of regular attendance.

Figure 1.

Student Attendance and Proficiency Rate  
in English, Language, and Arts

Data from Oregon Education Department



Interpret boxplots here

Figure 2.

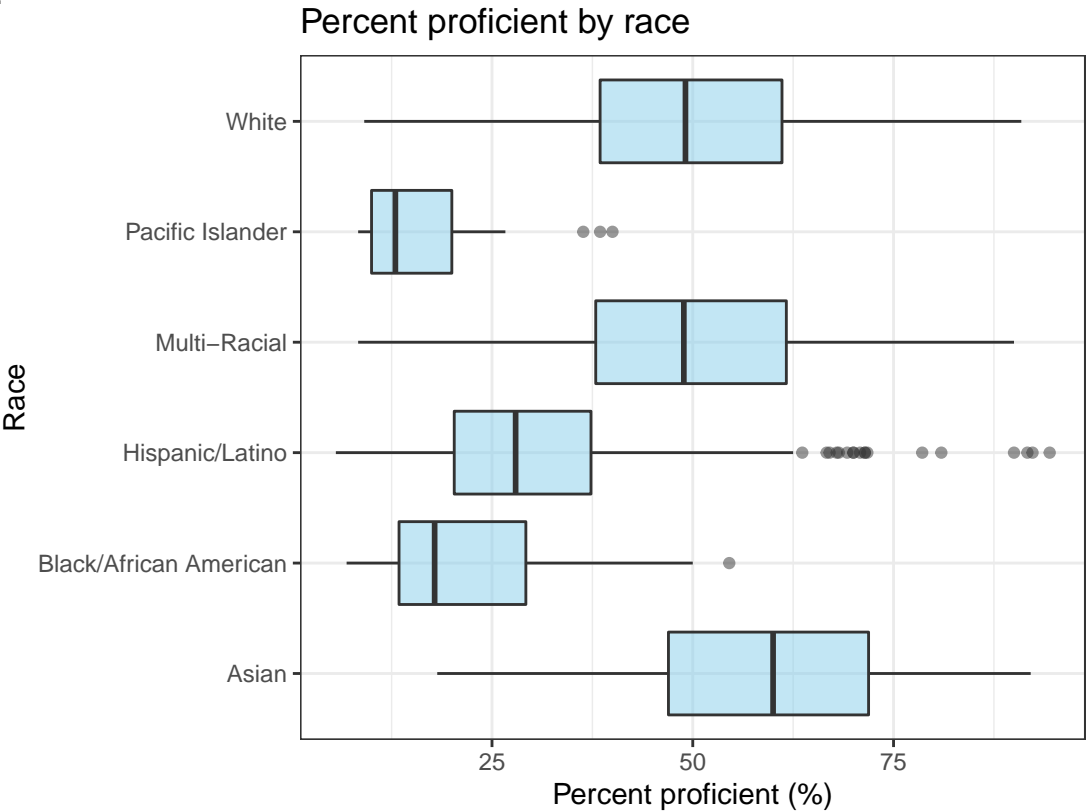
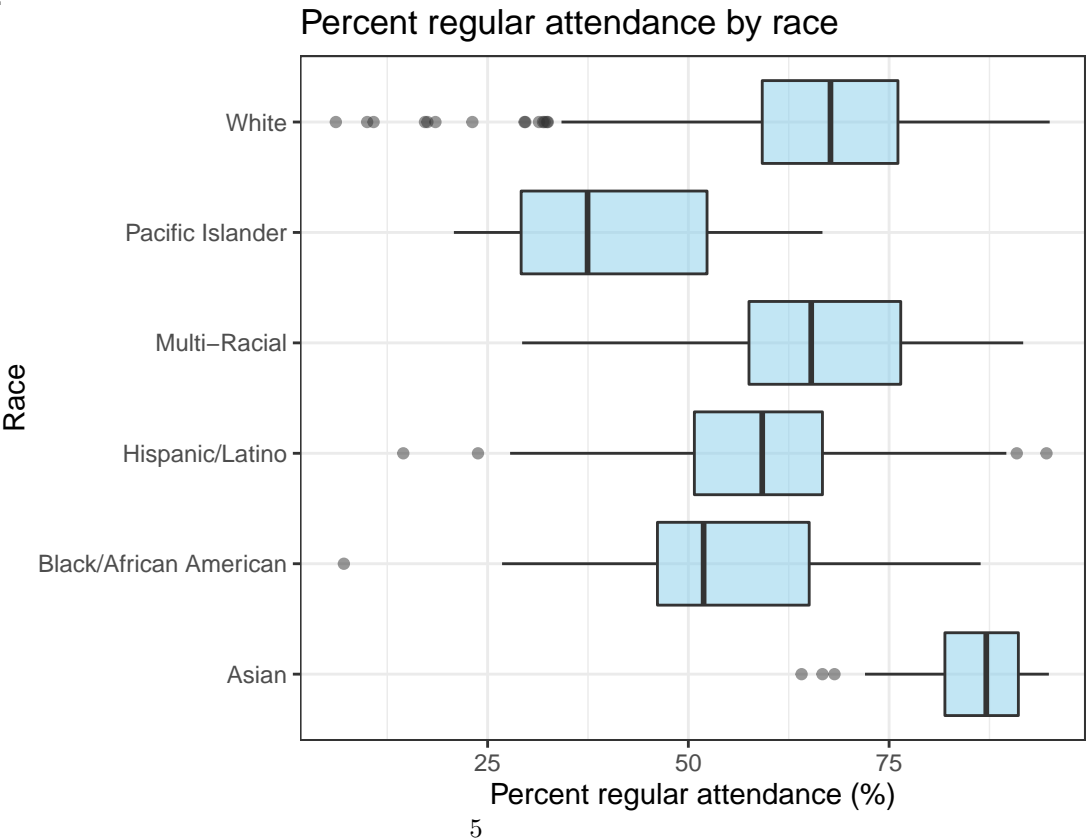
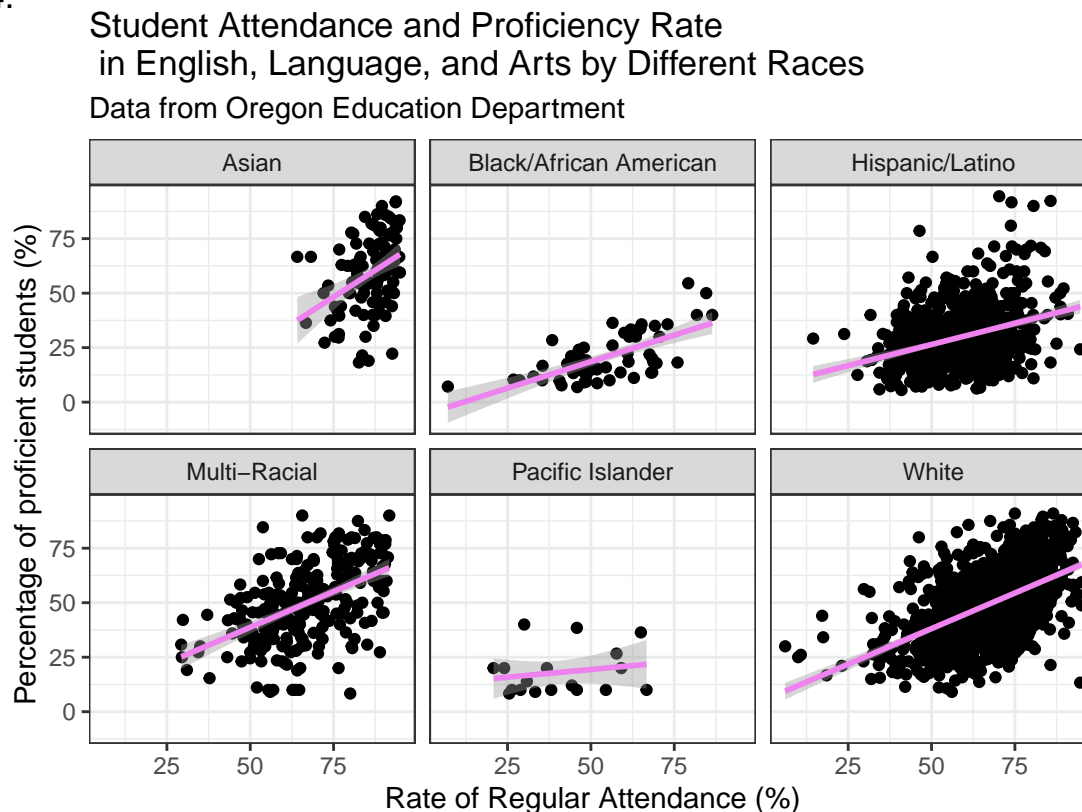


Figure 3.



In Figure 4 below, we used *facet\_wrap* function of {ggplot2} plot the rate of regular attendance by percentage of proficient students based on the race. To answer the second research question, we only included a visual representation of the data, but we didn't fit a regression model. Therefore, we evaluated the data only through a visual assessment.

Figure 4.



## Discussion & Conclusions

### 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.

To explain the varying effect of attendance rate on student proficiency across racial and ethnic groups, we suppose that there are other variables that correlate with racial and ethnic groups and it is those omitted variables, not racial and ethnic groups that have impacts on the relationship between attendance and

proficiency. For example, students coming from lower-income households that frequently experience food insecurity might have anxiety and stress that affect their performance, even when their attendance rates are the same as the reference group. This interpretation suggests that this research could be expanded to include other variables such as household incomes, free school lunch, and parents' education attainment.

Interestingly, all the data points for the Asian group show up 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 the lowest attendance - below 70% across all schools.

## Conclusion - (Merge this in the previous section)

Based on data of Oregon schools in the 2021-2022 academic year, our model predicts that for every 1% increase in attendance rate, schools will see an additional 0.75% in their percentage of students that are proficient in English, Language, and Arts, and the relationship is statistically significant. Our model suggest that a reasonable measure schools can use to improve proficiency rate is to encourage students to attend class more regularly. Policies such as offering free school lunch, addressing bullying, and providing teachers with diversity, equity, and inclusion training have the potential to improve attendance, and should be considered as schools attempt to improve student proficiency.

Our research also preliminary looked at how the relationship between attendance and proficiency varies across student racial and ethnic groups. While our visual exploration suggests the relationship does vary, additional research is needed to confirm this hypothesis and identify the relationship patterns.

## References

- Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Garcia, Emma, and Elaine Weiss. 2018. “Student Absenteeism: Who Misses School and How Missing School Matters for Performance.” *Economic Policy Institute*.
- Henderson, Tia, Caitlin Hill, and Kerry Norton. 2014. “The Connection Between Missing School and Health: A Review of Chronic Absenteeism.” *Portland, OR: Upstream Public Health*.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- Oregon Department of Education. 2022. “2020-21 Regular Attenders.” <https://www.oregon.gov/ode/schools-and-districts/reportcards/reportcards/Pages/Regular-Attenders-2021.aspx> <https://www.oregon.gov/ode/students-and-family/healthsafety/Pages/COVID19.aspx>.
- . n.d. “Reports & Data : State of Oregon.” Accessed December 6, 2022. <https://www.oregon.gov/ode/reports-and-data/pages/default.aspx>.
- Patnode, Amber Humm, Kim Gibbons, and Rebecca Edmunds. 2018. “Attendance and Chronic Absenteeism: Literature Review.” *Central for Applied Research and Educational Improvement*, 1–55.
- 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 Golemund, 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/>.