Impact of STEM Learning and Parental Education on Academic Achievement of High School Students

Alice (Shuhan Ai)

Graduate School of Education (HED)

Introduction

Student’s academic performance continues to be a top priority for educators. STEM course engagement, parent education level, family SES, student gender, and many other factors influence students' academic achievement both inside and outside of school.

One of the “hottest” skills in the 21st century workplace is the ability to use technology to solve complex and multidisciplinary problems. It has previously been observed that STEM education can serve as a means of integrating different knowledge and skills to prepare students to be an informed member of society (Honey et al., 2014). A growing body of literature recognizes that in order to promote 21st-century skills, students must have sufficient knowledge of STEM disciplines and STEM integration approaches, such as engineering design (Pleasants et al., 2020), computational thinking (Kong et al., 2020), and project-based and inquiry-based learning strategies (Brand, 2020). Furthermore, several studies have investigated factors that influence students' engagement in STEM fields. Sloanki and Xu (2018), for example, investigate the impact of instructor gender on student motivation in STEM fields. What's more, previous research has discovered that students' learning experiences in STEM courses influence their interest and success in subsequent learning, as well as their future academic decisions, such as major choice and college persistence (Chambliss & Takacs, 2014). Previous research suggests that participation in STEM courses improves students' overall academic performance and future development.

As schools and institutions work to increase access and quality of STEM education, other underlying factors that influence student academic achievement should not be ignored. Since the seminal 1966 Coleman Report, which highlighted family background as a critical component of educational production, academics from a variety of fields have sought empirical evidence of whether parental education and family socioeconomic status influence adolescent accomplishment.

According to Prof Charles Desforges (2003), parental participation is a crucial role in student accomplishment, and the more educated the mother, the greater the extent of involvement. According to Davis-Kean (2005), parental education is linked to their linguistic competence, and this has a substantial influence on how parents communicate with their children. Previous research indicates that the influence of parental education is substantial and is mediated by other variables, such as the availability of books at home, early literacy activities, and emergent literacy abilities at the time of school start (Myrberg, Rosen, 2010). Bourdieu (1990, 2002) defined a school diploma as a universally recognized symbolic capital that may explain unequal academic achievement of students from different social classes by connecting academic success to the allocation of cultural capital among different classes.

In this paper, I examine the impact of credits earned in STEM courses and parent's highest level of education on student's high school GPA using data from the High School Longitudinal Study of 2009 (HSLS:09). My goal is to comprehend the relationship between student academic achievement and STEM learning, as well as the influence of parental education on student academic achievement. In addition, I investigate a model for predicting student academic achievement using parental education and STEM learning. My research is guided by the following research question:

Is there an interaction relationship between the parent's highest level of education and student's STEM credits?

Hypotheses

According to the literature reviewed above, there is a positive association between student engagement in STEM courses and overall academic achievement. According to research, college-educated parents are more aware of the long-term benefits of obtaining a college degree and they will discuss this issue with their children. And parents who have obtained a higher degree are more likely to instill in their children the perseverance and abilities to navigate the path to success. I hypothesize that students whose parents have a higher degree will have higher GPA. Furthermore, I believe there is an interaction between the parental education degree and the student's STEM credits. As students take more STEM courses, the beneficial association between parental education and high school GPA will weaken. My hypotheses can be written, succinctly, as follow:

1. Assuming an interaction exists between STEM credits and parental education, students whose parents have higher degree levels will get higher GPA, and the trend will mitigate as they take more STEM courses.
2. Assuming an interaction dose not exist, students whose parents have higher degree levels will get higher GPA than those parental education level are less than high school.

Since this is a hierarchical hypothesis, my statistical analysis is in accordance with the principal of marginality. Namely, if the interaction is not significant, main effects should not be interpreted. Statistically:

H_0: $\beta_{interaction} = 0$. The coefficient for the interaction of parental education and STEM credits is 0.

H1: $\beta_{interaction} \neq 0$, $p < 0.05$. The coefficient for the interaction of parental education and STEM credits is not 0 with $p<0.05$.

Dataset

Data Source

This study utilizes data from HSLS:09, a nationally representative data set sponsored by the National Center for Educational Statistics (NCES) (Ingels et al., 2013). The HSLS:09 base year occurred in the 2009-2010 school year, when students were in ninth grade; the first follow-up occurred in spring 2012, when the majority of sample members were in eleventh grade. After students graduated from high school, a postsecondary update was conducted in the summer and fall of 2013; transcript data were also collected in 2013-2014. HSLS used a stratified, two-stage random sample design, with schools serving as primary sampling units in the first stage and students being chosen at random in the second (Ingels et al., 2013).

Variables

This study takes into account three variables: total high school GPA (X3GPATOT), STEM credits earned (X3CREDSTEM), and parent's highest level of education (X1PAR1EDU). I use student's high school GPA (X3GPATOT) to represent academic achievement as assessed by high school transcripts (2013-2014). The highest GPA achieved is 4.0. X3CREDSTEM denotes the total number of Carnegie credits earned in STEM courses, which uses the first two digits of the course SCED code: 02, 03, 10, 21. A Carnegie unit is also equivalent to a one-year academic course taken once a day, five days a week. The maximum STEM credits are 18. X1PAR1EDU indicates "parent #1's" highest level of education, as determined by the base year parent questionnaire. It is divided into seven categories: less than high school, high school diploma or GED, Associate's degree, Bachelor's degree, Master's degree, Ph.D/M.D/Law/other high level prof degree, and missing/non-response. In this study, I remove the missing data and non-respondents from the sample, only remain 6 levels.

Limitation

When preparing the HSLS:09 data set for estimation, it contains 23,503 samples and 9615 variables. The majority of these variables are not relevant for the present analysis. Here I will only focus on the three variables: student high school GPA, parental highest education level, and STEM credits. One of the assumptions of the linear regression is that model should include the functional form of the relationship that the variables are properly specified, and all relevant variables are included in the model. Since I only pay attention to three variables in this model, this could cause some limitation. After limiting the sample to individuals with complete information for the relevant variables, the sample for this study has 15,789 observations. I utilize the remaining 15,789 data to estimate the impact of STEM learning and parental education on high school students' academic achievement.

Methods

My hypotheses deal with differences between the levels of categorical variable. Therefore, my analysis is constructed to test for and ascribe magnitude and meaning to any potential difference. I plan to run three different statistical test: ANOVA, a post hoc Tukey test, and ANCOVA/multiple regression.

I include the ANOVA test to see if there is an interaction between parental education and the amount of STEM course credits earned by students. This question will help me decide which model to employ later. I will use type II sum of squares to test my interaction first, because my null hypothesis here is there is no significant interaction between parental education and STEM course credits. If there is indeed no interaction, then type II is statistically more powerful than other types. To account for the multiplicity issue, I perform a post hoc Tukey's test to confirm where the difference between parental education groups occurred.

If the ANOVA findings demonstrate that there is no association between parental education and STEM credits earned by the student, I will use the ANCOVA model to predict the student's overall high school GPA. If there is a significant association between parental education and student STEM credits earned, the ANCOVA model is invalid since it requires the two variables to be independent and unrelated. At this point, I'll introduce the interaction variable and build a multiple linear regression model to predict the student's overall high school GPA.

At the beginning of my analysis, I use basic descriptive statistics to understand the distribution of student academic achievement, parental education level, and STEM credits earned by students. The reason to present the descriptive and summary table is to test the basic assumptions of the linear regression model, including linear relationships, the random sample of observations, no multicollinearity, independence of error, and normality of the residuals. And if we violate any one of these assumptions, the model is uncertain. After analysis of graphics, summary statistics, I conducted my statistical analysis.

Findings

<u>*Descriptive statistics*</u>

Table 1

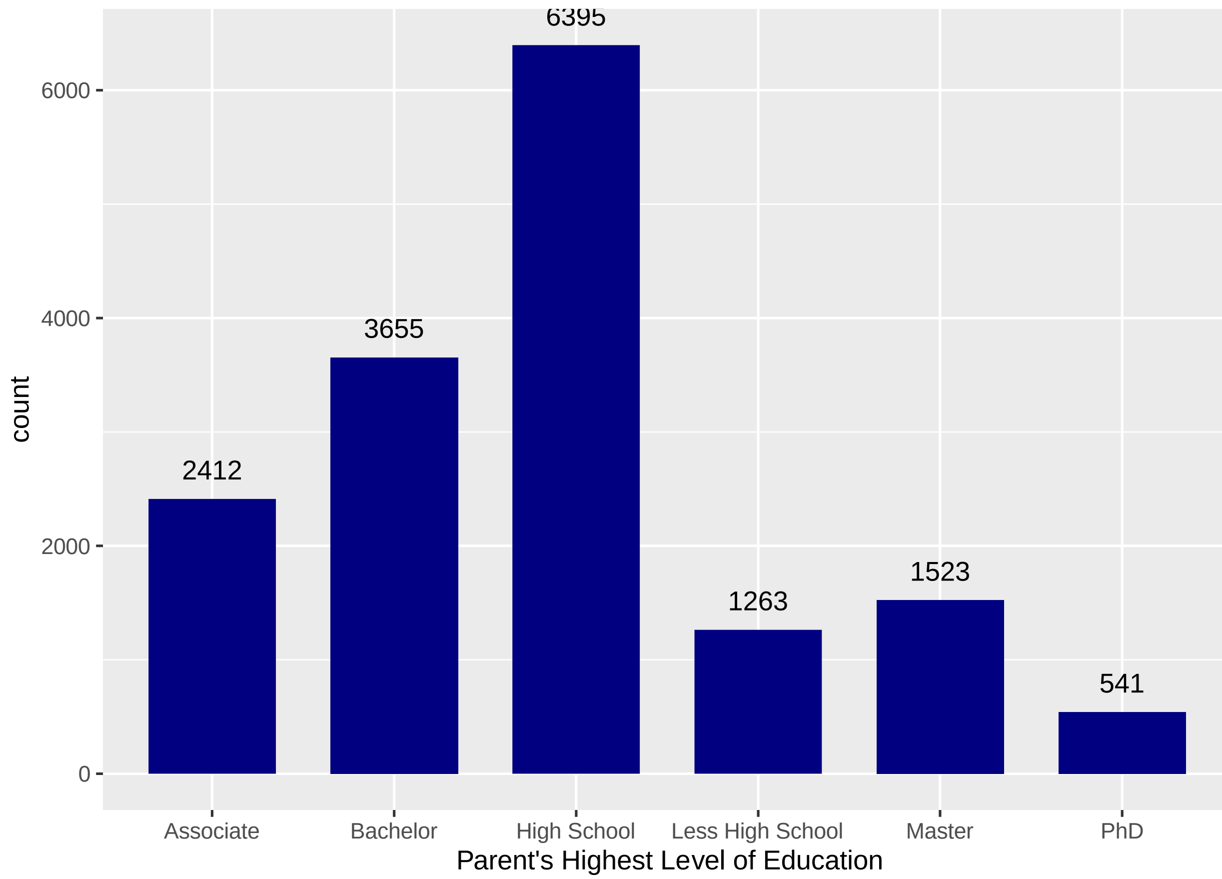
Descriptive Statistics of Student Total High School GPA and Credits Earned in STEM

Variables	numbers	mean	SD	min	max	skew	kurtosis
Total GPA	15789	2.81	0.83	0.3	4	-0.65	-0.02
Credits Earned in STEM	15789	7.70	2.54	0	16	-0.47	1.13

Table 1 shows the descriptive statistics for students' cumulative high school GPA and credits earned in STEM. The average overall high school GPA is 2.81, with a range of 0.3 to 4.0. The distribution is moderately tilted to the left, indicating that while some students have poor GPAs, the majority of students have high GPAs. The average number of credits gained in STEM courses is 7.70, with a range of 0 to 16. The skew indicates that some students do not take or take only a few STEM courses, whilst the majority of students take more STEM courses than the average number. The standard deviations of total GPA and credits earned in STEM are 0.83 and 2.54 respectively.

Figure 1

The Distribution of Parent's Highest Level of Education



The distribution of parents' greatest level of education is depicted in Figure 1. In this study, 6395 parents had a high school degree, accounting for 40.5 % sample. 36.2 % have a bachelor's degree or above, including 3655 with a bachelor's degree, 1523 with a master's degree, and 541 with a Phd/higher degree. There are 63.8 % of parents who have less than a bachelor's degree, including 6395 with a high school/GED degree, 2412 with an associate's degree, and 1263 with less than a high school diploma.

Figure 2 *The Regression of Student Total GPA and Credits Earned in STEM Grouped by Parent's Highest Education Level*

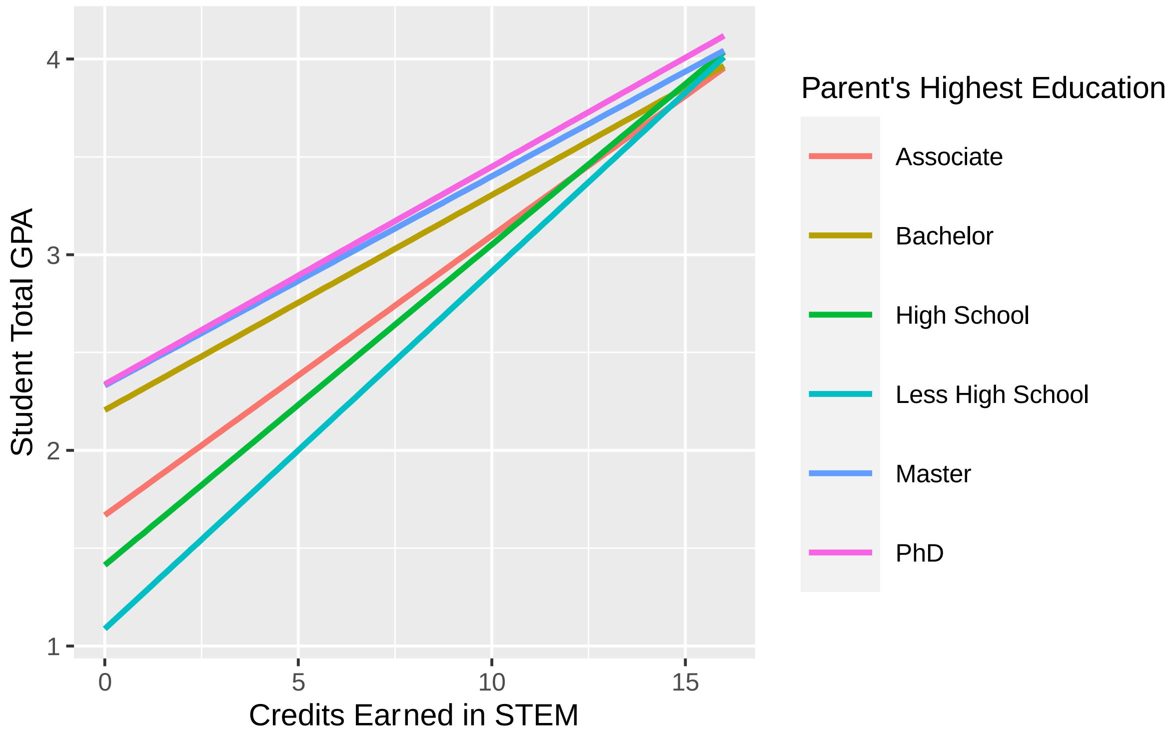


Figure 2 shows regression lines predicting number of placements given a child's age, grouped by race. Slight decreases in the slope of some lines can be perceived. For example, the slope of students whose parent are less than high school line seems to be higher than the rest.

ANOVA test, post hoc

Table 2 *Two-way ANOVA table with Credits Earned in STEM and Parent's Highest Education as IVs and Student's GPA as Dependent Variable*

Source	Df	SS	F-value	P-value	ETA^2	Partial ETA^2
Credits Earned in STEM	1	2811.6	6100.5	<2e-16 ***	0.260	0.279
Parent's Highest Education	5	649.5	281.84	<2e-16 ***	0.060	0.082
Credits Earned in STEM : Parent's Highest Education	5	66.0	28.66	<2e-16 ***	0.006	0.009
Residuals	15777	7271.410				

Table 2 shows a two-way ANOVA test with Credits Earned in STEM and Parent's Highest Education as IVs and Student's GPA as Dependent Variable. Notice in Table 2 that the p values all less than 0.05 indicate that there are significant effects for credits earned in STEM, parent's highest education, and their interaction. Starting with credits earned in STEM, the ETA^2 values is 0.26, indicating that 26% of the variance is accounted for by student's credits earned in STEM, whereas parent's highest education accounts for 6%, the interaction accounts for 0.6%. From the above analysis, we can conclude that parent's highest degree of education has a significant correlation with a student's STEM credits. Therefore, I use a multiple linear regression as my model to predict the student's overall high school GPA.

Table 3
Tukey Simultaneous Tests for Differences of Means

Contrast	Estimate	Difference of Means	DF	P-value
Associate - Bachelor	-0.28	0.0181	15777	<.001
Associate - High School	0.09	0.0163	15777	<.001
Associate - Less High School	0.27	0.0249	15777	<.001
Associate - Master	-0.38	0.0229	15777	<.001
Associate - PhD	-0.42	0.0350	15777	<.001
Bachelor - High School	0.38	0.0145	15777	<.001
Bachelor - Less High School	0.56	0.0237	15777	<.001
Bachelor - Master	-0.10	0.0217	15777	<.001
Bachelor - PhD	-0.14	0.0342	15777	<.001
High School - Less High School	0.18	0.0224	15777	<.001
High School - Master	-0.48	0.0202	15777	<.001
High School - PhD	-0.52	0.0332	15777	<.001
Less High School - Master	-0.66	0.0276	15777	<.001
Less High School - PhD	-0.70	0.0382	15777	<.001
Master - PhD	-0.04	0.0370	15777	0.896

Table 3 shows the results of the pairwise comparison. In almost every case, students whose parents are less than high school degree are expected to experience statistically significantly lower GPA than other students.

Regression Model and Goodness of Fit

Model: $y_i=\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\ldots+\beta_{interaction} x_{ip}+\epsilon$

The equation above describes the multiple regression of a student's overall GPA on STEM course credits, parental education, and the interaction between STEM credits and parental education. y_i represents student's high school GPA. b0 is an intercept when all of the independent variables are equal to zero. And the linear regression yields the results in Table 4.

Table 4
The Coefficients of the Model

	Estimate	**S.E.**	**T-value**	**Pr(>|t|)**	**UCI**	**LCI**
(Intercept)	1.840	0.026	69.365	< 0.01***	1.788	1.892
STEM Credit	0.136	0.003	44.138	< 0.01***	0.130	0.142
PhD	-0.206	0.055	-3.755	< 0.01***	-0.314	-0.098
Master	0.413	0.056	7.266	< 0.01***	0.302	0.525
Bachelor	-0.506	0.054	-9.320	< 0.01***	-0.612	-0.399
Associate	-1.247	0.082	-15.118	< 0.01***	-1.408	-1.084
High School	-0.006	0.136	-0.041	0.967	-0.272	0.261
STEM Credit:PhD	0.008	0.007	1.185	0.236	-0.005	0.021
STEM Credit:Master	-0.031	0.006	-4.749	< 0.01***	-0.044	-0.018
STEM Credit:Bachelor	0.030	0.006	4.717	< 0.01***	0.018	0.043

STEM Credit:Associate	0.074	0.010	7.323	< 0.01***	0.053	0.093
STEM Credit:High School	-0.004	0.015	-0.287	0.774	-0.034	0.025

Figure 3

Diagnostic Plots for the Linear Regression

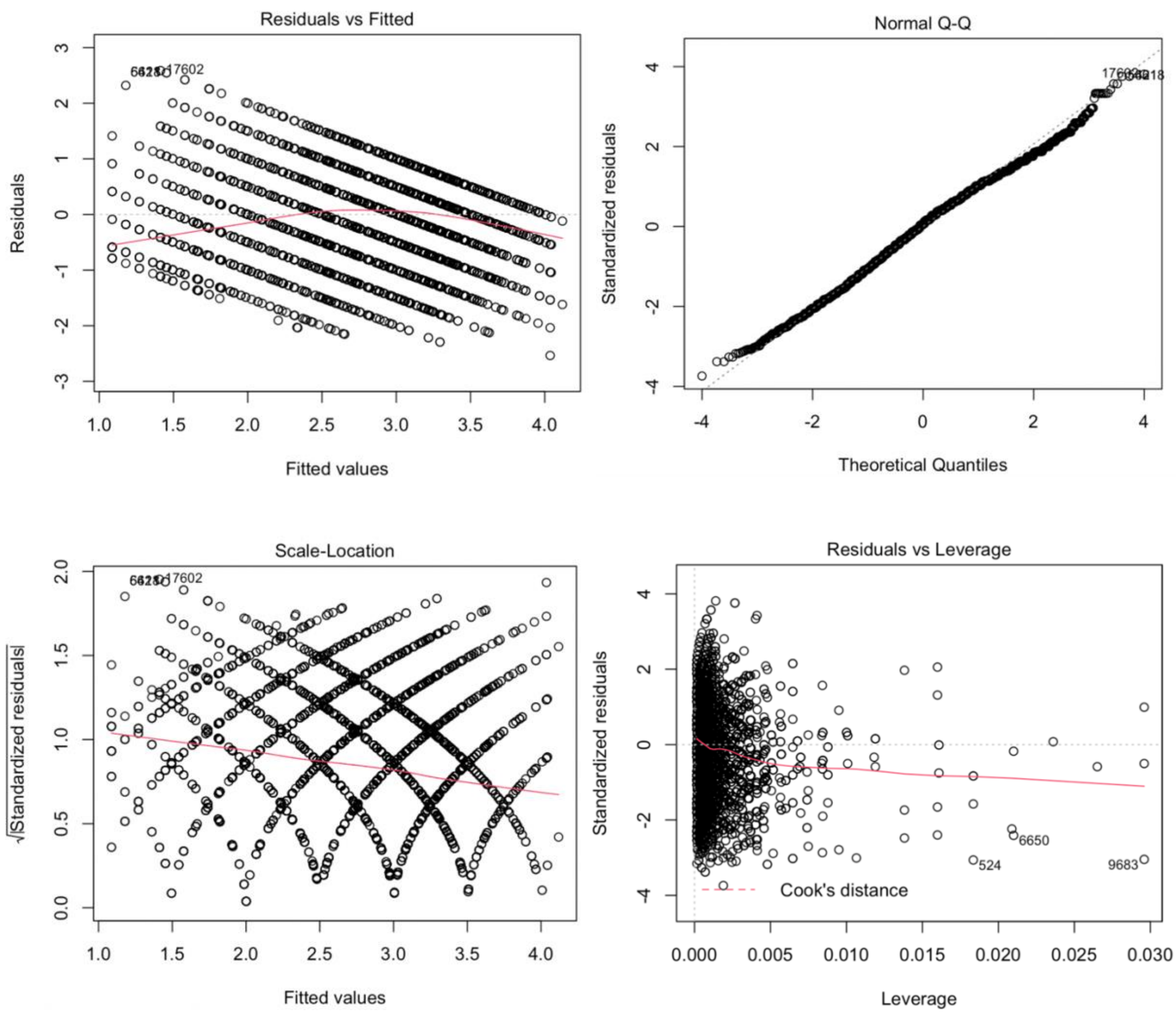


Figure 3 contains four diagnostic plots for the linear regression. The sum of the evidence from these four graphs is that this data is not well modeled linearly to some extent and that there are several influential points which may be influencing the outcome of tests. Normal Q-Q plot shows whether residuals are normally distributed. Although some observations look a little off, the overall distribution won't concern me because other observations still roughly distributed along this straight line. The Scale-Location plot shows that residuals are not spread equally along the ranges of predictors, indicating that the linear model does not fit the data well.

Limitations

This data analysis has several limitations. First, the linear model does not fit the data well, indicating that the reported size of differences is likely not accurate. Second, the model lacked sufficient factors that could reflect parental education and student STEM learning. The model's R2 is 0.33, indicating that the factors in the model can only explain 33% of the student's GPA. In a future study, researchers should include more variables such as parental education, family socioeconomic status, and student attitudes toward STEM learning to improve the model's explanatory power. Finally, the data is not random. This presents a potential problem for errors especially, where clustering could easily occur.

Discussion and Implications

This study emphasizes STEM learning and parental education as crucial variables in students' academic achievement. According to the research, students who take STEM courses may improve their academic achievement. The findings also indicate that parental education level has a positive impact on students' academic achievement.

Furthermore, evidence suggests that when students take a sufficient number of STEM courses, the effect of parental education level on student academic achievement might be somewhat offset. As a result, we should encourage disadvantaged students to enroll in STEM courses in order to spark their interest in study and boost their academic achievement.

Schools must begin to identify and encourage students to actively participate in STEM learnings, widen their horizons with rich practical STEM experiences, and develop their multiple thinking and creativity in STEM (Cheng et al., 2020). Educators should talk to students about using a STEM education to build a promising future, regardless of their age, sex, gender, and background. Previous research also indicates that high school students may become passionate about a STEM career if they have opportunity to find new role models within in STEM fields (González-Pérez et al, 2020). High school teachers could introduce students to incredible STME icons like Katherine Johnson, Peter Higgs to arise their interests in STEM learning. Finally, for policy makers, they could establish STEM scholarships for disadvantaged students and cover their tuition costs. Policy makers could also provide financial support to increase funding for STEM development in specific schools.

Conclusion

The data suggest that students whose parents are less than high school degree tend to have higher GPA when they take more STEM courses. The evidence points to the need for increases efforts to encourage students to learn STEM courses. And special attention should be ought to be paid to the disadvantaged populations.

Reference

Bourdieu, P., & Passeron, J. C. (1990). *Reproduction in education, society and culture* (Vol. 4). Sage.

Bourdieu, P. The forms of capital A. H. Halsey, H. Lauder, P. Brown, A. Wells Stuart *Education – culture economy, society* 46– 58 Oxford Oxford University Press 2002.

Brand, B. R. (2020). Integrating science and engineering practices: outcomes from a collaborative professional development. *International Journal of STEM Education*, 7(1), 1-13.

Chambliss, D. F., & Takacs, C. G. (2014). How college works. In *How College Works*. Harvard University Press.

Cheng, Y. C., & So, W. W. M. (2020). Managing STEM learning: a typology and four models of integration. *International Journal of Educational Management*.

Crawford, S. L. (2006). Correlation and regression. *Circulation*, *114*(19), 2083-2088.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. *Journal of family psychology*, *19*(2), 294.

Desforges, C., & Abouchaar, A. (2003). *The impact of parental involvement, parental support and family education on pupil achievement and adjustment: A literature review* (Vol. 433). London: DfES.

González-Pérez, S., Mateos de Cabo, R., & Sáinz, M. (2020). Girls in STEM: Is it a female role-model thing?. *Frontiers in psychology*, 11, 2204.

Honey, M., Pearson, G., & Schweingruber, H. (2014). *STEM Integration in K-12 Education: Status, Prospects, and an Agenda for Research.* Washington, DC: The National Academies Press.

Ingels, Steven J., Daniel J. Pratt, Deborah R. Herget, Jill A. Dever, Laura Burns Fritch, Randolph Ottem, James E. Rogers, Sami Kitmitto, and Steve Leinwand. 2013. High School Longitudinal Study of 2009 (HSLS:09) Base Year to First Follow-Up Data File Documentation. Appendixes. NCES 2014-361. Washington, DC: National Center for Education Statistics.

Kong, S. C., Lai, M., & Sun, D. (2020). Teacher development in computational thinking: Design and learning outcomes of programming concepts, practices and pedagogy. *Computers & Education*, 151, 103872.

Myrberg, E., & Rosén, M. (2009). Direct and indirect effects of parents' education on reading achievement among third graders in Sweden. *British Journal of Educational Psychology*, *79*(4), 695-711.

Pleasants, J., Olson, J. K., & Cruz, I. D. L. (2020). Accuracy of Elementary Teachers' Representations of the Projects and Processes of Engineering: Results of a Professional Development Program. *Journal of Science Teacher Education*, 31(4), 362-383.

Appendix R Code

```
#Descriptive Statistics
knitr::kable (describe (df_final ), digits = 2)

#parents education plot
plot_hist <- ggplot(df_final, aes(x=factor(f_X1PAR1EDU)))+
  geom_bar(width=0.7, fill = "navy")+
  geom_text(stat='count', aes(label=..count..), vjust=-1)
theme_minimal()
print(plot_hist +
        labs(y = "count", x =" Parent's Highest Level of Education"))

#classify by parents' education
plotedu <- ggplot( df_final, aes(X3TCREDSTEM, X3TGPATOT, color=X1PAR1EDU)) +
  geom_smooth(method =lm, se = F, fullrange =T) +
  scale_color_discrete(name = "Parent's Highest Education") +
  theme(legend.key.size = unit(1, "cm"))
print(plotedu +
        ggtitle("The Regression of Student Total GPA And Credits Earned in STEM Grouped by
Parent's Highest Education")+
        labs(y= "Student Total GPA", x= "Credits Earned in STEM"))

#Anova test for the model type
Fit1 <- aov(X3TGPATOT ~ X3TCREDSTEM*f_X1PAR1EDU, data=df_final)
summary(Fit1)
anova_stats(Fit1)
Anova(ancova, type="2")

#Tukey
Tukey <- lsmeans(Fit1, pairwise ~ f_X1PAR1EDU:X3TCREDSTEM, adjust = "tukey" )
Tukey

#helmert coding
my.helmert = matrix(c(5/6, -1/6, -1/6, -1/6, -1/6, -1/6,
                      0, 4/5, -1/5, -1/5, -1/5, -1/5,
                      0, 0, 3/4, -1/4,-1/4,-1/4,
                      0, 0, 0, 2/3, -1/3, -1/3,
                      0, 0, 0, 0, 1/2, -1/2), ncol = 5)
my.helmert

contrasts(df_final$f_X1PAR1EDU) = my.helmert

Fit <- lm(X3TGPATOT ~ X3TCREDSTEM*f_X1PAR1EDU, data=df_final)
summary(Fit)
confint(Fit, level=0.95)
plot(Fit)
```