Jaimie Choi

Economics 388

Library's Impact on Educational Development

Introduction.

Currently in the United States, there are approximately 119,487 libraries. Libraries are gaining a larger presence on Americans' lives. According to American Library Association, libraries are not only seen as places with books but are also "viewed as anchors, centers for academic life and research and cherished spaces". Over ten years, there was a 20 percent increase in library visitations and 54.4 percent increase in library program attendance (American Library Association, 2015). My research paper will specifically analyze whether increase in library funding affects academic achievements of children in the United States.

The United States, a country with the "most technologically powerful economy" and the third largest GDP, struggles with a high child poverty rate (CIA, 2016). The child poverty rate in the United States was 20% in 2014 and 21% in 2015 (Kids Count Data Center, 2016). In fact, the high child poverty rate seems to be a significant reason why the United States scored "lower than the average for all OECD countries" in math and not as well in reading as compared to economically robust countries (National Center for Educational Statistics). Krashen (2011) argues that American students have not done well in international tests when compared to other countries because of child poverty. In fact, he states that "studies show… that American students from well-funded schools who come from high-income families outscore nearly all other countries on those kinds of tests" (17). He claims that the high child poverty rate brought down the national average test scores significantly (Krashen, 2011). Payne and Biddle (1999) explain that the poor children "are uniquely handicapped" because of poverty since they have "little

access to books, writing materials, computers, and other supports for education that are normally present in middle class or affluent homes in America." (7).

A very convenient tool for students of low-income families to use to perform well academically is the public library. In a public library, they can freely borrow books, acquire computer skills, participate in reading classes, and take different classes such as English Learning and Career Exploration (American Library Association, 2015). In fact, the analysis of the data from "representative national samples in 27 nations, with over 70,000 cases" shows that children in homes with "many books get three years more schooling than children from bookless home, independent of their parents' education, occupation, and class" (Evans, 2010). An effective way for children to have access to books and technology at no cost to themselves or their parents is to improve and increase the amount of libraries. The American Library Association explains this issue by specifying that one challenge for them is the limited service of many "small and rural libraries" which negatively affects the poor Americans in those areas (American Library Association, 2015). Achterman (2008) gives convincing evidence of the importance of libraries by studying the effect of California school library's collection size and hours open on California standardized test scores. Through regression analysis, he found that an increase in library factors causes small but significant increases in English language art test score and social studies test scores. (Achterman, 2008).

Currently there seems to be no current analysis of the effect of libraries on students across the United States. Most studies focus more on specific small areas, different countries, or specific types of libraries or programs. This paper will look at general data of United States libraries and schools from the past few years to determine if increasing investment on libraries will have some significant impact on children's educational levels. If we find that increased

expenditure on library has little effect, then perhaps a change in the use of the investment money is necessary.

Data.

My research project will use data from 50 states at five different time periods. Because of limited data, I can only analyze the effect from 2003 to 2013, with a year gap in between each year. I found many data on libraries in the Public Library Service Data from the Institute of Museum and Library Services. This government funded institution has been surveying US public libraries annually since 1988. I found the total expenditure on libraries and divided it by the population of the state to more effectively compare larger and smaller states with each other. In addition, I collected data from this government service on the number of kid books circulated and kid program attendance to determine if expenditures led to an increased investment on children education in the library.

The National Center for Education Statistics provides the average test scores for different topics of the National Assessment of Educational Progress test. This is a "largest nationally representative and continuing assessment of what America's students know and can do in various subject areas" (National Center for Education Statistics). I used eighth grade test scores because they are a good medium between scores of elementary students and high school students. I chose to use math scores, along with reading scores, to see if better libraries help students become more academically well-rounded. Greater use of library books or attendance in library programs could help students think more critically, thus helping students in any academic subject.

I found other controlling variables to prevent an upper bias in the effect of library expenditures on average test scores. I used data on annual personal income per capita from Bureau of Economic Analysis because, as stated by Krashen (2011), family income is very correlated with the child's education level. I found educational attainment percentages of people 25 years or older from the United States Census Bureau. I specifically found the percentage of those who received a high school diploma or higher and the percentage of those who received a bachelor's degree or higher to balance my econometric model. This data accounts for the fact that parents' education level does impact their children's academic achievements. I also added data on state unemployment rates from Bureau of Labor Statistics to account for children struggling academically due to parents struggling to find jobs to support them. These data sources are all from the United States bureaucratic organizations. I trust these transparent government databases to be unbiased and factual. In addition, I also included the state child poverty rate for children under 18 years old in my model to compare state economic disadvantage level. Child poverty is defined as children living with families with incomes below the federal poverty level. This data is collected by National Kids Count with a purpose to "track well-being of children in the United States." This is from a non-profit organization that compiled and edited data from large US Census surveys.

Here is a chart of the summary statistics of the variables used in this research.

Table 1 Summary Statistics of the Various Independent Variables

Variable	Observation	Mean	Standard Deviation	Minimum	Maximum
Total Expenditure Per Capita	300	31.87302	11.5094	13.0904	60.99375
Kid Book Circulation	300	1.55e+07	1.61e+07	1319654	9.24e+07
Kid Program Attendance	300	1191205	1163279	50548	6436861

Eighth Grade Scores	300	545.231	13.38303	512.5306	578.0095
Eighth Grade Reading Scores	300	263.821	6.090008	248.5054	277.0096
Eighth Grade Math Scores	300	281.41	7.697146	261	301
Income Per Capita	300	37587.43	7536.099	22550	64110
Child Poverty Percentage	300	0.1868	0.0527291	0.08	0.34
Percentage of High School Diploma and Higher	300	86.50933	3.633877	77.2	92.7
Percentage of Bachelor Degree and Higher	300	26.96433	4.810582	15.1	39.1
Unemployment Rate	300	6.349333	2.144613	2.6	13.6

In this summary statistic, it appears that there is a significant standard deviation of total expenditure per capita per state and test scores for different years. This could perhaps indicate possible correlation between library expenditure and eighth grade test scores.

Research Design.

To analyze the most accurate effect of library expenditures on student test scores, I made regression equations more extensive to cover all the assumptions to have an unbiased coefficient estimator.

First, I started with simple regressions that examined whether increasing library expenditures improves the attendance of children programs and circulation of children books. Here are the regression models I used:

$$KidBookCirculation_{state,\ year} = \beta_0 + \beta_1 TotalExpenditurePerCapita_{state,\ year} + \varepsilon_{state,\ year}$$

$$KidsProgramAttendance_{state,\ year} = \beta_0 + \beta_1 TotalExpenditurePerCapita_{state,\ year} + \varepsilon_{state,\ year}$$

Next I regressed eighth grade scores on total expenditure per capita. This regression answers the research question in the simplest way:

$$EighthGradeScores_{state,\ year} = \beta_0 + \beta_1 TotalExpenditurePerCapita_{state,\ year} + \varepsilon_{state,\ year}$$

The simplicity of the model sheds light on many of its problems. This model has omitted variable bias because there are many other factors that influence a student's test scores. There needs to be controlling variables so that the upper bias of the coefficient of total expenditure per capita is minimized. In addition, heteroscedasticity and serial correlation is highly likely to occur because this data is collected from all over the United States and from different time periods. Heteroscedasticity and serial correlation will produce incorrect standard errors, and thus false confidence intervals.

To fix omitted variable bias, I ran a multivariate regression that included different variables mentioned in the data section. I used the following model below:

$$\begin{aligned} EighthGradeScores_{state,\ year} &= \beta_0 + \beta_1 TotalExpenditurePerCapita_{state,\ year} + \beta_2 IncomePerCapita \\ &+ \beta_3 ChildPovertyPercentage + \beta_4 UnemploymentRate + \beta_5 Highschooldiploma \\ &+ \beta_6 bachelorsdegree + \varepsilon_{state,\ year} \end{aligned}$$

In order for this model to provide accurate predictions of the effect of these various variables, it needs to adhere to assumptions to have a consistent estimator. I will show tests in the following section that prove that heteroscedasticity is present. In addition, endogeneity still exists. The total expenditure per capita most likely has a lag of 2 years because expenditures after the standardized test is taken will not affect the test in any manner.

To fix omitted variable bias, I ran a regression using fixed effects and lag variables. Fixed effect fixes autocorrelation by taking out variables in the error term that affect outcome, such as race and gender. In addition, fixed effect with clustered panel variables gives you robust standard errors that solve heteroskedastic issues.

Finally, I ran more specific regressions with fixed effects, clustered variables, and lag variables to see whether total library expenditure affects one subject score more than the other. The next section shows the results and tests regarding these different regression models.

Empirical Evidence.

I will show the different results of regressions starting from a simple bivariate model to a more complex, in-depth model appropriate for panel data sets. I will report results of hypothesis tests to show that the end model fits the assumptions necessary for the estimator to be unbiased.

Before I begin with the result of linear bivariate regression, I will show the effect of total expenditures on the number of kid books circulated and kid program attendance per year.

First, I did two regressions of the number of kid books circulated and kid program attendance on total expenditure. I put the regression under a Breush-Pagan Test to detect possible heteroskedasticity. This is so that when I do the hypothesis test to see if total expenditures affects the amount of kid books circulated or kids program attendance, the test statistics will produce accurate results. The Breush-Pagan test gave large chi-squared values of 472.03 and 164.50 with p values close to zero. This means that there is significant evidence that heteroscedasticity exists. Thus, I ran the two regressions with robust standard errors. This means that there is significant evidence that heteroscedasticity exists.

Regression with robust standard errors:

1. Regression of kidbooks circulation on total expenditure

Table 2 Estimate of the Effect of Library Expenditures on Amount of Kids Books Circulated

Independent Variables	Amount of Kid Books Circulated	Robust Standard Error
Total Expenditure	0.0633***	0.0031829
Constant	2,822,167***	467,269
Observations	300	
R-squared	0.8902	

*** p<0.01, ** p<0.05, * p<0.1

2. Regression of kids' program attendance on total expenditure:

Table 3 Estimate of the Effect of Library Expenditures on Attendance in Kid Programs

Independent Variables	Number of Children who attended Kid Programs	Robust Standard Error
Total Expenditure	0.00456***	0.0001761
Constant	279,611.7***	29,995.16
Observations	300	
R-squared	0.8852	

*** p<0.01, ** p<0.05, * p<0.1

According to the charts, total expenditure has a greater impact on the circulation of kid books than the kids program attendance. Yet, the coefficients are very small, meaning an increase of library expenditure will only cause a miniscule change in the circulation of children books and children program attendance. Nonetheless, since the test statistics for total expenditure on both variables are fairly large and the p-values are essentially 0, there is significant evidence that the coefficients do not equal zero. Thus, increase in total expenditure most assuredly results in an increase in children book circulation and program attendance.

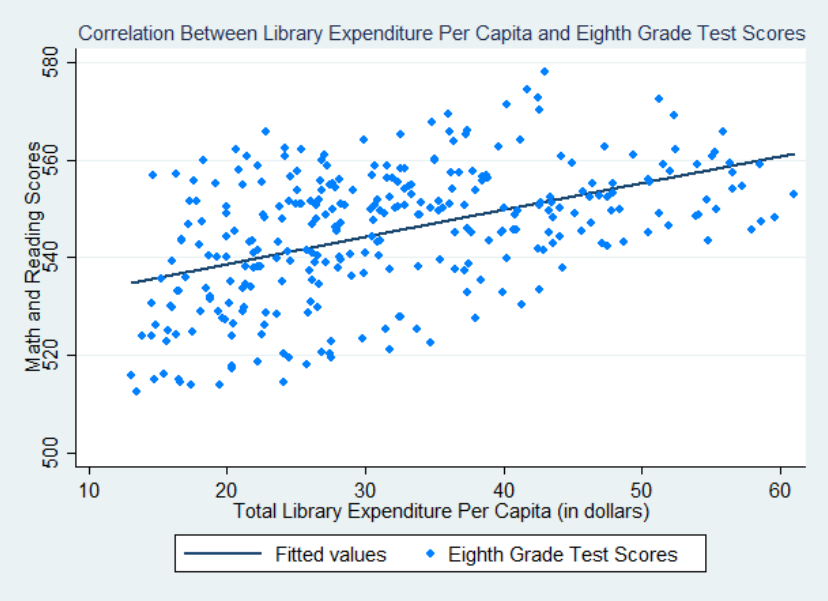
Consider the simple linear model chart that shows the regression of eighth grade scores on total expenditure per capita:

Table 4 Bivariate Estimate of the Effect of Library Expenditures on Eighth Grade Test Scores

Independent Variables	Eighth Grade Test Scores	Standard Errors
Total Expenditure Per Capita	0.5508***	0.05932
Intercept	527.67***	2.0099
Observations	300	
R-squared	0.2244	

*** p<0.01, ** p<0.05, * p<0.1

This scatter plot graph displays the linear regression result:



In this simple regression model, the coefficient for total expenditure per capita is large enough to show its strong influence on test scores. In fact, with the large t statistic of 9.29 and p-value of virtually 0, there is significant evidence to reject that total expenditure per capita has no

effect on test scores. However, many assumptions are violated when using this bivariate linear model. There is a substantial upper-bias on the coefficient of total expenditure per capita.

In order to account for missing variables in the simple bivariate model, I ran a multivariate linear regression:

Table 5 Multivariate Estimates of the Effect of Library Expenditures on Eighth Grade Test Scores

Independent Variables	Eighth Grade Scores	Robust Standard Errors	95% Confidence Interval
Total Expenditure Per Capita	0.0175	0.0614	(-0.1034, 0.1384)
Percentage of High School Diploma and Higher	2.0822***	0.2126	(1.664, 2.5005)
Percentage of Bachelor Degree and Higher	0.362**	0.1595	(.0481, .6759)
Child Poverty Percentage	15.2446	18.69	(-21.5475, 52.0366)
Income Per Capita	0.000505***	0.0000946	(.00032, .00069)
Unemployment Rate	0.0901	0.3045	(-.5092, .6894)
Intercept	332.408***	20.76	(291.4996, 373.3153)
Observations	300		
R-squared	0.5866		

*** p<0.01, ** p<0.05, * p<0.1

The difference between the coefficient of total expenditure per capita from simple bivariate regression and that of the multivariate regression is 0.5333, a large difference. This indicates that the omitted variables caused a large over exaggeration of library expenditure's effect on students' test scores. Percentage of high school diploma and higher and income per capita seem to have the most significant effect on test scores. However, the other dependent variables may have correlation issues with each other. A simple regression of eighth grade test

scores on child poverty percentage suggested a strong inverse relationship. However, here an increase in child poverty causes a significant jump in the test scores. In addition, the test statistic for total expenditure per capita is very low and the p-value is high. In this case, I don't have enough evidence to reject the case that there is no effect of total expenditure per capita on test scores. To check if the test statistics are accurate, I did a Breusch-Pagan Test and Modified White Test to check for heteroscedasticity. According to Breusch-Pagan Test, the chi-squared value is 3.60 and the p-value is 0.0576. There seemed to be heteroscedasticity but there was no significant evidence. As a result, I did the modified White Test to ensure if heteroscedasticity will be an issue. According to this test, the chi-squared value is 67.06 and the p-value is 0. There is substantial evidence that heteroscedasticity exists.

Finally, I ran the Hausman Test to check for endogeneity by testing fixed effect and random effect. This informs us whether fixed effect or random effect will be the best way to fix the failed assumptions. For the Hausman Test, the chi-squared result is 85.52 and the p-value is 0. Given the very low p-value, there is significant evidence of endogeneity and it is necessary to use fixed effect on the regression.

To fix heteroscedasticity and serial correlation, I ran a regression model with a lag variable for total expenditure per capita, fixed effect, and cluster variables. Observe the regression results below:

Table 6 OLS Estimates of the Effect of Library Expenditures on Eighth Grade Test Scores

Independent Variables	Eighth Grade Test Scores	Robust Standard Errors	95% Confidence Intervals
Total Expenditure Per Capita	-0.1737	0.1716	(-0.5184, 0.1711)
Total Expenditure Per Capita Lagged 2 years	0.1726	0.139	(-0.1068, 0.4521)

Percentage of High School Degree and Higher	0.1539	0.3111	(-0.4713, 0.7791)
Percentage of Bachelor Degree and Higher	-1.1061***	0.3185	(-1.7462, -0.4661)
Child Poverty Percentage	51.5514***	18.2043	(14.9685, 88.1343)
Income Per Capita	0.00058***	0.000174	(0.00023, 0.00093)
Unemployment Rate	0.724***	0.1606	(0.3865, 1.032)
Intercept	525.9681***	23.4453	(478.853, 573.083)
Observations	250		
Number of states	50		
R-squared	0.673		

*** p<0.01, ** p<0.05, * p<0.1

The coefficient of total expenditure per capita is very intriguing. Contrary to our assumption, total expenditure over capital seems to have little to no influence on the eighth grade test scores. The coefficient is -0.1737 for total expenditure per capita and 0.1726 for the lagged version of it. For both independent variables, the confidence intervals go from a negative coefficient to a positive coefficient. When running the hypothesis test, both total expenditure per capita and the lagged version of it have very small t values and large p-value. There is no sufficient evidence against the hypothesis that total expenditure per capita has no effect on the scores. The R-squared value increased from 0.5866 to 0.673 when comparing the fixed effect regression to the simple multivariate regression. This suggests a greater correlation between the variables and test scores.

Here is a regression that replaces total scores with just reading scores: regression that replaces total scores with just reading scores:

Table 7 OLS Estimates of the Effect of Library Expenditures on Eighth Grade Reading Scores

Independent Variables	Eighth Grade Reading Scores	Robust Standard Errors	95% Confidence Intervals

Total Expenditure Per Capita	-0.1887**	0.0851	(-0.3597,-0.0178)
Total Expenditure Per Capita Lagged 2 years	0.2028***	0.071	(0.0601, 0.3455)
Percentage of High School Degree and Higher	-0.1349	0.163	(-0.4626, 0.1927)
Percentage of Bachelor Degree and Higher	-0.6578***	0.1697	(-0.9999,-0.3168)
Child Poverty Percentage	40.4484***	10.0569	(20.2384, 60.6584)
Income Per Capita	0.000176**	0.000081	(0.0000134, 0.000339)
Unemployment Rate	0.216**	0.0913	(0.0326, 0.3994)
Intercept	277.34***	12.3499	(252.5205, 302.1567)
Observations	250		
Number of states	50		
R-squared	0.587		

*** p<0.01, ** p<0.05, * p<0.1

Notice that the coefficient of the total expenditure per capita is 0.1887 and its lagged version's coefficient is 0.2028. There is a 0.015 decrease in the coefficient for total expenditure per capita and a 0.0302 increase in the coefficient for the lagged variable. In this case, the test statistic of the lagged variable is larger than the original fixed effect regression and the p-value is significantly smaller. Here, I reject the null hypothesis that total expenditure per capita lagged does not have an effect on eighth grade reading scores. Total expenditure per capita that is lagged for 2 years has a significant positive effect on students' reading scores.

Here is the regression of all the variables on eighth grade math scores:

Table 8 OLS Estimates of the Effect of Library Expenditures on Eighth Grade Math Scores

Independent Variables	Eighth Grade Math Scores	Robust Standard Errors	95% Confidence Intervals
Total Expenditure Per Capita	0.0151	0.1015	(-0.1888, 0.219)
Total Expenditure Per Capita	-0.0302	0.0909	(-0.2129, 0.1525)

Lagged 2 years			
Percentage of High School Degree and Higher	0.2888	0.1932	(-0.0994, 0.6770)
Percentage of Bachelor Degree and Higher	-0.4484**	0.2149	(-0.8803,-0.0164)
Child Poverty Percentage	11.103	10.363	(-9.7221, 31.9282)
Income Per Capita	0.000404***	0.000095	(0.0002, 0.0006)
Unemployment Rate	0.4933***	0.0981	(0.2961, 0.6905)
Constant	248.63***	16.23	(217.838, 279.420)
Observations	250		
Number of states	50		
R-squared	0.633		

*** p<0.01, ** p<0.05, * p<0.1

The coefficients of the total expenditure per Capita is 0.0151 while the lagged version is -.0302. The test statistic values are small and the p-values increased significantly. This gives inadequate evidence to reject the hypothesis that these variables have no effect on math scores. Total expenditure per capita doesn't seem to impact math scores.

lt seems that remaining endogeneity issues have caused the secondary variables to have distorted effects on test scores. The regression results all estimate that the percentage of having a bachelor's degree or higher has a negative correlation with eighth grade test scores. This seems counterintuitive because parents with higher educational level tend to have better careers and also give motivation to the children to attain at least the same degree. Fortunately, I focused mainly on the effect of library funding on test scores instead of seeing what kind of effect other variables have on scores. However, perhaps adding instrumental variables in the future will prevent the secondary independent variables from skewing the coefficients of each other. Also, the lack of many years prevent a more accurate estimate of the effect of library spending on academic achievement by eliminating short-run trends.

Conclusion.

Results show that increase in total library expenditure has little to no influence on the overall eighth grade test scores. However, the results give strong evidence that library expenditure affects eighth grade reading scores. This makes intuitive sense because libraries focus more on reading comprehension than developing math skills. However, it seems that the critical reading skills do not transfer well to math for the students. In order to see how libraries could help students get higher math grades, there could be research done using panel data on a few specific libraries with different services. Then, a regression analysis can be made to see what service is more effective. For example, the paper could focus on comparing libraries from cities of similar size and demographics to see if there are differences after funding after-school math programs. Also, in order to increase the effect of libraries on reading scores, libraries could use more of the funding to help entice more students to read books by advertisement and reward programs. This may increase the coefficient of total library expenditure when eighth grade reading score is regressed on it.

Further studies on library impact on education will help libraries improve and provide more effective free services to people who especially need them. For example, studies that analyze the effect of library funding on science test scores could also shed further light on the effectiveness of libraries on students. I believe there is huge potential in using libraries to alleviate poverty in the United States and end poverty cycles that last for generations. My hope is that libraries can one day bridge the gap between students' performance from low income families and students' performance from those of high income.

References

Achterman, Douglas L. *Haves, Halves, and Have-Nots: School Libraries and Student Achievement in California*. Denton, Texas. UNT Digital Library, Dec 2015. Web. 8 Dec. 2016.

"American Fact Finder." *United States Census Bureau*. U.S. Department of Commerce, n.d. Web. 16 Nov. 2016.

"Children in Poverty (100 Percent Poverty)." *Kids Count Data Center*. The Annie E. Casey Foundation, n.d. Web. 16 Nov. 2016.

"Contact CIA." Central Intelligence Agency. Central Intelligence Agency, 2016. Web. 07 Dec. 2016.

"International Comparison of Achievement." *National Center for Education Statistics*. N.p., n.d. Web. 07 Dec. 2016.

Krashen, Stephen. "Protecting Students Against the Effects of Poverty: Libraries." *New England Reading Association Journal* 46.2 (2011): 17-21. Web. 16 Nov. 2016.

"Local Area Unemployment Statistics." *Bureau of Labor Statistics*. U.S. Bureau of Labor Statistics, 26 Feb. 2016. Web. 16 Nov. 2016. Multiple webpages for Different Years

M.D.R. Evans, Jonathan Kelley, Joanna Sikora, and Donald J. Treiman. "Family Scholarly Culture and Educational Success: Books and Schooling in 27 Nations." *Research in Social Stratification and Mobility* 28.2 (2010): 171-97. Web. 16 Nov. 2016.

"NAEP State Comparisons." *National Center for Education Statistics*. N.p., n.d. Web. 16 Nov. 2016.

"News Release Archive." *Bureau of Economic Analysis: U.S. Department of Commerce*. Bureau of Economic Analysis, 3 Dec. 2016. Web. 3 Dec. 2016.

Payne, Kevin J., and Bruce J. Biddle. "Poor School Funding, Child Poverty, and Mathematics Achievement." *Educational Researcher* 28.6 (1999): 4-13. Web. 16 Nov. 2016.

"PLS Data and Reports." *Institute of Museum and Library Services*. Institute of Museum and Library Services, n.d. Web. 16 Nov. 2016.

"Public Library Use." *American Library Association*. American Library Association, Oct. 2015. Web. 16 Nov. 2016.