Impact of Total Per Student School Expenditure on Intergenerational Upward Mobility

Tate Clemen
Drake University

Subject Area: Economics
Article Type: Editorially-Reviewed Journal Article

ABSTRACT

Looking at the effects of per student school expenditures on intergenerational upward mobility, this research examines if there is a significant positive effect from increased school spending on upward mobility of children. Data from more than 10 million children across 720 Commuting Zones was used to analyze this effect. The results of the study align with previous literature that show there are other factors that hold a greater effect on mobility than school spending. This research also shows that increased school spending has a greater, positive effect on those in the lowest tail of the income distribution than those in a higher area of the distribution.

INTRODUCTION

The ‘American Dream’ has no one clear definition, but one way to look at it may be the ability for a child to out-earn their parents. With the classic view of this ‘Dream’, and the United States (U.S.) being classically hailed as the ‘land of opportunity,’ one might hope that there would be the opportunity available for an individual to become even more successful than the generations that came before them. Interested in the systems that cause inequality, this paper will delve into inequity in education and its effect on the achievement of the so-called American Dream.

Currently in the United States, the public school system is set up in such a way that tends to allocate resources to those that are likely already resource-rich. Public schools are funded in a variety of different manners, but for a majority of states, almost half of total revenues come from property tax. In 2016-2017, some 82 percent of local revenues for public schools came from property taxes (National Center for Education Statistics, 2020). With this, we would suspect that there would be massive disparities between those on the upper and lower halves of the income distribution, as those in the upper-half likely reside in property-rich areas, while those in property-poor areas are likely in the bottom tail of the income distribution. This system of fund distribution can be compared against the Nordic education system whose policies equalizes funding across public schools.

While there are many possibilities of what could contributes to the later-life achievement gaps we see between students, access to resources and spending have been brought to the table as a driving factor in performance inequalities.

Building on the work of Chetty, Hendren, Kline, and Saez (2014) on socioeconomic mobility, this paper will attempt to determine the magnitude and significance of the effect of per student school expenditures on intergenerational upward mobility. While there are copious amounts of research on the effects of education and the effects of school spending on children in the long run – this literature will be reviewed in the subsequent section – this paper looks specifically at the effects of per-pupil school spending on the relationship between a child's future earnings as compared to their parents' earnings. Following the literature review, Section 2, there will follow a section on the data used for this research. The data explanation (Section 3) will be followed by Section 4, coverage of the methodologies used, followed up by the results in Section 5, and the research will be concluded in Section 6.

LITERATURE REVIEW

The impacts of education have been well researched within the economic community, and under that umbrella, there has been plenty of research looking at school funding specifically. The following section of this paper will describe the earlier research on returns to schooling, school funding, long-term student performance, and intergenerational mobility.

Theoretically, as resources increase for students, we would expect to see increases in later-life success: success, for the purposes of this research, will be considered monetarily. There is a long line of literature that ties K-12 academic success to increased future earnings (Chetty, et al., 2011; Watts, 2020, Currie & Thomas, 2001). All of this literature, along with plenty of others, points to a strong positive relationship between test scores, among other factors, in K-12 education and future earning and employment. This meaning, students with higher test scores seemed to achieve higher salaries later in life.
From here, we look toward what may be causing these higher test scores. Greenwald, et al. (1996) looked at the effect of school resources on student performance, and they deduced that there may be a significant increase in performance from only a moderate increase in spending. The most significant factors that they found for student performance was per-pupil expenditure (PPE), student-teacher ratio, teacher experience, and teacher salary. This research seems to hold up against other literature, but there is also plenty of research that would claim different.

On the other side of this argument, Dearden, et al. (2002) found that after controling for ability and family background, the student-to-teacher ratio had no effect on education qualifications and men's wages, but it did have an effect on women's wages, especially those with lower ability levels. Much like this, Chetty, et al. (2011) found that classroom size has little effect on later-life income, but their research did show that students in smaller classrooms had a significantly higher chance of attending college and exhibited notable increases in improvement on other outcomes.

With all of this, it would make sense that we would see higher rates of upward mobility given increases in school spending. Chetty, et al. (2014) – from which this research is inspired – found that the geographic areas with the highest rates of mobility had (1) less residential segregation, (2) less income inequality, (3) better primary schools, (4) greater social capital, and (5) greater family stability. The research in this paper differs in the sense that it will focus exclusively on the impact that schooling has on mobility. With Chetty, et al. (2014) finding that better primary schools is one of the leading factors in mobility, it would be a logical next step to see if increased spending – which would likely occur in theses 'better schools' – would also have a significant positive effect on upward mobility.

There has been plenty of research done on the the effects of schooling on later-life monetary success. With that, there has been quite a bit of research done on the correlation of per-pupil expenditures on educational outcomes, but there has only been a minimal amount of work that looks directly at the relationship between school expenditures and mobility.

DATA

In order to test whether there is a significant effect on upward mobility from an increase in school spending, data was gathered on mobility and school expenditures, along with data for the various controls.

The data and measures of intergenerational mobility that are used to conduct this research were assembled by Chetty, et al. (2014) and was obtained from Opportunity Insights' online data library. This data looks at multiple cohorts of children born from 1980-1985. For this project, these cohorts were averaged across years to create a cross-section. This change was made in an effort to create a cross-sectional dataset for each Commuting Zone (CZ), rather than the panel that Chetty had used.

All of this mobility data stems from federal income tax records from 1996-2012, and further links them to earlier records connecting the individual children in the sample to their parents. All of these children had to have (1) a Social Security Number (SSN) or an Individual Taxpayer Identification Number, (2) a birth date between 1980 and 1991, and (3) had to have been US citizen as of 2013. This creates a total sample of over 10 million children.

These individuals are all broken up into their respective Commuting Zones (CZs). These Commuting Zones were established by Tolbert & Sizer (1996) and are constructed by combining various counties. In determining which counties to combine, Tolbert & Sizer (1996) looked at commuting patterns found from the 1990 Census. These helped to include the rural areas that surround the more heavily trafficked, typically more urban regions. In total, there are 741 commuting zones comprised of all 3,006 counties in the U.S. These CZs are what are used as the observations for this research, resulting in 741 total observations, with none left out.

As mentioned earlier, the 1980-1985 birth cohorts were averaged to create a singular observation group. This was done in an effort to alleviate any singularity or selection bias within a particular birth year. This cohort averaging also creates the opportunity to look at the dataset as a cross-section, rather than looking at it as a panel across the different cohorts.

In **Table 1** below, the key variables used in this research are listed, along with their definition.

Table 1. Variable Definitions

Variable	**Definition**
Average Absolute Mobility	The expected rank of children whose parents are at the 25th percentile of the national income distribution, averaged across 1980-1985 cohorts
Avg. Absolute Mobility at the 50^{th} Percentile	Average absolute mobility across 1980-1985 birth cohorts for the 50^{th} percentile of parental income. Calculated by taking AM + (50-25)*RM
School Expenditure Per Student	Average expenditures per student in public schools (1996-1997 school year) (Reported in 000s)
Urban Area (dummy)	Dummy variable indicating whether a CZ is an urban area, determined by whether a CZ intersects with a metropolitan statistical area (MSA)
Household Income per capita	Aggregate household income in the 2000 census divided by the number of people aged 16-64. (Reported in 000s)
Gini	Gini coefficient computed using parents of children in the core sample, with income top-coded at $100 million in 2012 dollars

Sources: Chetty, et al. (2014), George Bush Report Card, NCES CCD 1996-1997 Universe Survey, 1992 Census of Government county-level summaries, NCES CCD 2000-2001, Tax Records

---

The aforementioned Absolute Mobility variable was initially calculated in Chetty, et al. (2014) on an individual level within each cohort. This variable was created through the use of a ‘rank-rank’ calculation. This calculation is based off of ranking the individual children in each respective cohort based on their adult income and then comparing that rank to the rank of those individual children’s parents’ incomes to the other parents’ incomes within that given cohort. This rank-rank metric identifies the correlation between parent income and adult child income. The absolute mobility metric that is used in this study is for those children whose parents fall within the 25^{th} percentile of the income distribution. In other words, those in the lower fourth of the income distribution. This mobility measure for Chetty’s three individual cohorts is what was averaged out to create the average absolute mobility metrics that will be used for this research.

In an effort to see if there are differing affects for those higher in the income distribution, a second dependent variable is going to be tested. This is the average mobility at the 50^{th} percentile. This variable was calculated using the Average Absolute Mobility (AM) variable from earlier plus the 50 (the percentile we are looking to get) minus 25, multiplied by the Relative Mobility (RM) metric. Relative Mobility is the slope from OLS regression of child rank on parent rank within each CZ. This variable was also averaged across the three cohorts to create a single Relative Mobility metric. Chetty, et al. (2014) indicated that the mobility rating for any other percentile (P) could be measured by taking AM+(P-25)*RM. So, the end equation to achieve the absolute mobility metric for the 50^{th} percentile was calculated by taking AM+(50-25)*RM for each respective commuting zone.

The school expenditure data, which will be our key independent variable, is from the 1996-1997 school year. The rationale behind this stems from an effort to incorporate a lag from the actual birth. This puts birth cohorts within the middle of the K-12 system, based on typical age. This means the funding would be when the cohort individuals are ages 12-17, in line with the typical enrollment ages for middle and high school students. This data was taken from the George Bush Report Card and was cleaned by Chetty, et al. (2014).

Below, in **Table 2**, the descriptive statistics of the variables used in this study are listed.

Table 2. Descriptive Statistics

	Number of observations	Mean	Standard deviation	Minimum	Maximum
Avg. Absolute Mobility	729	44.74043	5.829538	26.9315	64.66925
Avg. Absolute Mobility at the 50th Percentile	729	52.61157	4.859638	36.498	69.8715
School Expenditures per Student (1000s of dollars)	731	6.036724	1.18577	3.920151	11.90626
Urban Area	741	.4385965	.4965505	0	1
Household Income per capita (1000s of dollars)	741	32.86996	5.750707	16.69586	58.62839
Gini	741	.4055269	.0811415	.20178	.84733

Sources: Chetty, et al. (2014), George Bush Report Card, 2000 Census, Tax Records

These variables all seem like valid variables to control for when looking at the effect of school expenditures on intergenerational mobility. Average absolute mobility was chosen as our dependent variable and is an absolute measure of what we are treating, which makes it a perfect variable for the study. As stated previously, this measure of mobility was originally calculated in Chetty, et al. (2014). It appears that the descriptive statistics for both of our dependent variables make sense, since the expected ranks are in the 0-100 range. With this, we see that the range for the 50th percentile is higher, which would seem to make sense, since we would expect that the higher an individual's parental income, the higher income rank they are likely to reach in adulthood.

School expenditures per student is what is being used as the treatment variable, and it is measured on a per-pupil basis. This is done in an effort to equalize the measurement across schools, so there is no bias for size of educational institution or student population. This equalization should create more comparable data. The expenditure variable was divided by 1000 to achieve a more accessible metric. Looking at the summary statistics, it would make sense that per-pupil spending is between ~$4,000 and ~$12,000. This is also a pretty large range, which confirms the thinking that there are fairly large disparities in per-pupil school spending across different regions. With this, we are hoping to see how much of an, if any, impact this factor has on mobility measures.

The first control that is taken into account is urban area. This is a dummy variable that was assigned to the data by Chetty, et al. (2014) on a Commuting Zone level. In determining which observations were considered 'urban', we look toward regions denoted as metropolitan statistical areas (MSAs). An MSA, similar to a CZ, is the formal name of a region that consists of a city and surrounding regions, based off of social and economic factors. For this research, a CZ will be considered 'urban' if it intersects with an MSA. This variable is controlled for given the vast difference in resources for an urban area compared to a more rural area. This discrepancy could have major effects on mobility levels. Along with this, there could be discrepancies in school input costs. Monk (2007) noted that there are distinct salary differences between educators in urban versus nonurban areas, where typically teachers in urban areas, on average, receive a higher salary.

Following that, we add in household income per capita as a control. This helps to achieve a measure of wealth across a given CZ, and this is also a vital factor in the calculation of school funding. This is a precise wealth measurement for a respective area, given that we typically see an increase in resources and mobility given the amount of wealth in an area. Our descriptive statistic here again would seem to make sense. An average of $16,000 to $58,000 would make complete sense, since would reasonably fall in the average U.S. income range. Here again, it can be seen that there are massive disparities among these areas, and this would likely contribute to the mobility rate. This data was pulled from the 2000 U.S. Census and was calculated by taking the aggregate household income and dividing it among those ages 16-64. Again, this data was pulled and cleaned by Chetty, et al. (2014).

And finally, the income Gini index is taken into consideration. The Gini coefficient is the measure of inequality based on distribution of income. For this work, the Gini is based off parent family income, based on the core sample. This is done given the likelihood that if there is a larger income distribution in an area, an individual would have a farther range to go to rise in income rank. The numbers in the descriptive statistics make complete sense, given that the Gini can only range from 0-1, and we are not very likely to see perfect income equality within an area, which would create a Gini coefficient of 0, and we are also not likely to see complete income inequality within an area, which would be a coefficient of 1. That said, with a range from .2 to nearly .85 would suggest that there are areas with exceptionally high ranges of income distribution that will need to be controlled for. This variable was calculated by Chetty et al. (2014) from tax records.

METHODOLOGY

This research is attempting first to estimate the following equation:

$$\textit{Absolute Mobility at } 25^{th} \textit{ Percentile} = \beta 0 + \beta 1\ \textit{SchoolExpenditures} + \beta 2 \textit{UrbanArea} + \beta 3 \textit{Income} + \beta 4 \textit{Gini} + \varepsilon$$

The outcome variable here, as mentioned previously, is the absolute mobility rate at the 25^{th} percentile averaged across 1980-1985 birth cohorts. The treatment variable here is the total school expenditures per student. The control variables here are as follows: a dummy variable for urban area, household income per capita, and the Gini coefficient. These are used in an effort to account for other significant factors that may be at play when looking at education expenditures and mobility. The regressions that will be run will estimate the coefficients for these factors, represented in the equation by their respective beta (β).

While this equation may predict a correlation between school expenditures and intergenerational mobility, it is not likely that there this will be able to identify a causal effect. Rather, this research seeks to find the significance and magnitude of the effects of school expenditures on upward mobility. That said, it will be interesting to see if there is actually some level of correlation between the two variables. From a theoretical stand point, it would make sense that there is a positive correlation between increased school spending on a child and that child's ability to advance higher and achieve more in life.
To see if there are varying results for those in higher percentiles, this research is also planning on estimating the following equation:

$$\textit{Absolute Mobility at } 50^{th} \textit{ Percentile} = \beta 0 + \beta 1\ \textit{SchoolExpenditures} + \beta 2\textit{UrbanArea} + \beta 3\textit{Income} + \beta 4\textit{Gini} + \varepsilon$$

This equation will be estimated using the same variables as the previous regression, but the dependent variable will be changed to represent the Absolute Mobility at the 50^{th} percentile, as was mentioned previously. Looking at this higher percentile is important to gage whether the effects of school spending are felt throughout the income distribution or if there is a disproportionate effect on those in the lower fourth.
There is likely some level of selection bias, given variations in commuting zones. For example CZs with high levels of school spending are likely fundementally different from those CZs with lower school expenditures. These areas have differences in many other important ways that could be affecting mobility.

That said, there are undoubtably a number of other factors that could be affecting mobility that have not been factored into this equation. This research has not taken into account such factors as parental education, race, family stability, immigrant population, nearby colleges and universities, etc., all of which could have effects on mobility. Furthermore, this research will likely suffer from some level of omitted variable bias.

Each of these regressions should help to shed some light on the relationship between per-pupil expenditures and intergenerational upward mobility. In the results section to follow, we will see the kind of impact school spending has on mobility.

RESULTS

The results of all eight regressions are listed below in **Table 3**. The first four regressions shown use the main variable Average Absolute Mobility rate, which is taken at the 25^{th} percentile. In regressions 5 through 8, the Average Absolute Mobility at the 50^{th} percentile is used as the dependent variable. These two variables had interestingly different results.

Table 3. Estimates of the Effect of Per Student Expenditures on Intergenerational Mobility

	AM at 25th Percentile				**AM at 50th Percentile**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Per Student Expenditure (000)	1.233* (.165)	1.125* (.155)	.8844* (.177)	.3269 (.174)	.7761* (.139)	.6802* (.131)	.5484* (.145)	.0700 (.142)
Urban Area (dummy)		-4.485* (.371)	-5.127* (.418)	-3.313* (.384)		-3.976* (.307)	-4.328* (.355)	-2.771* (.307)
Household Income Per Capita (000)			.151* (.043)	.155* (.041)			.083* (.038)	.086* (.033)
Gini				-36.322* (2.875)				-31.178* (2.180)
Adjusted R-square	0.060	0.208	0.224	0.438	0.035	0.204	0.2108	0.441
Number of observations	720	720	720	720	720	720	720	720

Notes: Dependent variable for regressions 1-4 is Average Absolute Mobility for the 25th percentile. The dependent variable for Regressions 5-8 is Average Absolute Mobility at the 50th percentile. Standard errors robust to heteroskedasticity are shown in parentheses.
*significant at the 95th percentile

The first regression here shows that, without controlling for anything else, per student expenditures had a significant, positive effect on absolute upward mobility. This is saying that for every increase of $1,000 spent on a student, we would likely see a one-point increase in that child's expected future income rank. As predicted, there are positive effects across the board for our treatment variable. That said, what is interesting here is the difference seen between children at the 25th percentile and children at the 50th percentile. The regressions suggest that there is a more pronounced effect on the lower part of the income distribution from per student spending. It appears that as students rise in the income distribution, they experience diminished returns from increased school spending. Again, this would make sense, since those in higher ranking families are more likely to have a number of additional resources at home than those in the lower income range. That is to say that students from lower income homes are more likely to benefit from the increased funding at school.

The results from the addition of the urban dummy variable are interesting. Contrary to my earlier thoughts, it appears that being in an urban area tends to have a negative effect on mobility. This is consistent across all eight regressions and is constantly significant. While I thought that the additional resources available in urban areas would increase mobility, it appears that this feature has a negative effect. Looking at Chetty, et al. (2014), this appears to align with their findings. Also, this would make sense, since we are more likely to see larger income distributions in these areas, which will be touched on soon.

What seems to have a surprising low, but significant, coefficient is household income per capita. Across the board, household income did have a positive coefficient as suspected, but it did have a relatively low coefficient, lower even than the school expenditures. This would seem to say that average household income would not have a very large effect on mobility. This variable too seems to have a disproportionately larger effect on the lower end of the distribution compared at the 50^{th} percentile.

While the initial equation showed a fairly sizable effect on mobility from increased spending, this factor did decrease as more factors were controlled for. Once the Gini was added, the treatment variable become insignificant for both of the dependent variables. Along with this, the Gini also had the largest coefficient. The coefficient sizes make sense, since the Gini would be reported in a 0 to 1 measurement. This is estimating that a 0.10 increase in the Gini coefficient for an area would cause the expected rank of child in the 25^{th} percentile to decrease by 3.6322. The addition of the Gini also had the largest effect on the R-squared, adding over 20 points for each set of regressions. This aligns with the findings in Chetty, et al. (2014) which showed that income segregation – the uneven geographic distribution of income groups within a given area – was one of the top contributors to mobility rates. It also makes sense that the Gini has a negative coefficient, because as income distribution widens, it will become harder to move upward, given the range is so much larger.

CONCLUSIONS

This research set out to find the significance and magnitude of the impact of per student school expenditures on intergenerational upward mobility. This was done by looking at data that spanned multiple decades, covered 720 Commuting Zones, and represented more than 10 million U.S. children.

While the results do seem to say that there is a positive effect on mobility from school expenditures, it appears that there are other much larger forces at play. While not the main driver of mobility, it is interesting to see the difference in the effects of school spending at different income percentiles. It appears that there is a greater effect from increased school spending on those in the lower percentile than when those in higher income levels are added to the equation.

There were a number of limitations to this research. Given the premier data on mobility – that from Chetty et al. (2014) – is given at the Commuting Zone level, it is difficult to see potential disparities that could be within those respective areas. Multiple school districts may lie within a respective CZ, and these different districts may have distinctive budgets and/or spending patterns.

From a policy perspective, it appears that increased school spending should be focused more on students from the bottom tail of the income distribution. While total school spending itself cannot be directly targeted toward certain students, schools with high levels of student poverty may benefit from more revenue.

For future research, it may be valid to look into school funding equalization on intergenerational mobility. Since school funding is heavily based on local property tax, it may be interesting to see what the effects of implementing an equalization policy may have been. With this, it would also be interesting to see future literature on how the results of funding equalization efforts in the U.S. differ from those in other countries.

REFERENCES

Chetty, R., Freidman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. The Quarterly Journal of Economics, 126(4), 1593–1660.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. Quarterly Journal of Economics, 1553-1623.

Currie, J., & Thomas, D. (2001). Early Test Scores, School Quality and SES: Longrun Effects on Wage and Employment Outcomes. Research in Labor Economics, 35, 103–132.

Dearden, L., Ferri, J., & Meghir, C. (2002). The Effect of School Quality on Educational Attainment and Wages. The Review of Economics and Statistics, 84, 1-20.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The Effect of School Resources on Student Achievement. Review of Educational Research, 66(3), 361-396.

Monk, D. H. (2007). Recruiting and Retaining High-Quality Teachers in Rural Areas. Future Child, 17(1), 155-174.

National Center for Education Statistics. (2020). Public School Revenue Sources.
Tolbert, C. M., & Sizer, M. (1996). U.S. Commuting Zones and Labor Market Areas: A 1990 update. Economic Research Service Staff Paper, 9614.

Watts, T. W. (2020). Academic Achievement and Economic Attainment: Reexamining Associations Between Test Scores and Long-Run Earnings. American Education Research Association Open, 6(2), 1–16.

DATA APPENDIX

The data for this research was pulled from the Opportunity Insights' (opportunityinsights.org) data library (opportunityinsights.org/data), under the section for Chetty, et al. (2014), found here. For this research, the data was pulled on or around February 17, 2021.

The data was then taken from the online_data_tables.xls file and was pasted into a new excel file, which I entitled ECON_190_data.xlsx (also available upon request) in order to have a cleaner transfer over to Stata.

Once the tables were pasted into the new file, the row of numbered labels below each column's label was removed. This was done in an effort to not include those in an observation row.

After this step, each sheet in the Excel book was individually transferred into Stata to create a .dta file.

For one of the variable files, there were empty observations at the bottom of the data sheet, these had to be dropped so the files could merge.

Once these files were saved as .dta files, they were merged using the CZ ID into a single .dta file in Stata.

This file was the Stata .dta file that was used in the study.

Following this, the first variable that was generated was the average of the AMs for the respective cohorts. This was done using the egen function, and rowmean for the three cohorts.

This same function was done for the RMs for the three cohorts to create an average RM variable.

Then the AM variable at the 50^{th} percentile was generated using the equation AMavg50prct = AMavg + (50-25)*RMavg

Household income also needed to be divided by 1000 to create a Household income variable that is more manageable. This was done by generating HHIncomeperCap1000 = HouseholdIncomepercapita/1000

All of the regressions used regress function, with the vce (robust) function to provide robust standard errors.

Data and Stata code used to compute the estimates are available from the author on request.