Where Does the Pin Point on Educational Inequality in the United States

Lara L. Mechling

Bellevue University

DSC 101 Introduction to Data Science

Professor Neugebauer

October 10, 2021

Where Does the Pin Point on Educational Inequality in the United States

Project Problem

What are the main factors causing continued educational inequality in the United States?

Description

Children across the United States must attend school from Kindergarten through 12th grade, and schools are made available to these children at no cost to the parents. But, will each child be given the same opportunities and receive the same quality of education at every public school in the United States? Paulus, Spinath & Hahn describe educational inequality "to be present if students' educational outcomes are related to their social background or to economic and social circumstances over which students have no control" (2021). One major factor in educational inequality is the socioeconomic situation of parents. In Germany, "children with very high paternal occupation status were six times more likely to receive a recommendation for upper secondary school (access to university studies) compared to children with low parental occupational status" (Paulus, Spinath, & Hahn, 2021).

Parents send their children to schools in their communities based on government-issued zoning, which dictates where each child can attend school based on the location of the family residence. "The U.S. school system was founded on local funding and local control … and local funding remains a critical and continuous aspect of almost all state systems" (Card & Payne, 2000). Schools are funded through local taxes, with some state and federal assistance. Looking at the income disparity throughout the United States and even the income disparities inside cities, it is no wonder that schools receive different funding levels across the country. Per the data on educationdata.org, the state that spends the most money per student, at $23,321, is New York,

and the least money spent per student, at $7.610 per student, is Idaho (Hanson, 2021). These numbers say a lot about the disparity in educational funding under the current system.

Between the funding policies of the education system, the parent's socioeconomic status, and the child's ability, which plays the most significant factor in educational inequality? Do these factors work in tandem, and can a model be produced to predict how educational disparities could be improved using the data collected?

Project Scope

This project will attempt to determine if socioeconomic status, funding policies, and children's innate abilities are factors leading to educational inequality and look for any other attributes leading to educational inequality that may arise during the study. The project will use data to build a model showing how each characteristic affects the student. This model will be deployed to determine students and communities running the risk of receiving poor or unequal educations. The final goal of the study is to use the model to determine if there is a way to begin closing the educational gap by previewing the results of changes in areas of input (i.e., would a change in government funding change the success or failure rates of students in the area). Results of this study could influence government policy, assist non-profit organizations, and enlighten local school districts. The study will not predict a child's ability to be successful in school or their chosen career.

Data Sources and Elements

Table 1

Data Source, Data Element Selected from the Source, and Purpose for Use

Data Element	Data Purpose
District Cost Database	
district	Identify the school district.
state_name	Identify the state.

enroll	Total student enrollment.
year	Identify the year.
ppcstot	Actual spending per pupil. Here we will see how much each school district spends per student as calculated by the spending of the school districts in comparison to the number of students attending school in those districts. The data will allow insight into the amount of funds allotted in each school district for educational purposes.
predcost	Adequate spending per pupil. The School Finance Indicators Database (SFID) calculates an estimate of what each school district should be spending per student to achieve a common academic goal and it is derived from the National Education Cost Model (NECM) which is a part of the SFID databases. The data included in this element will allow insight into which school districts are spending funds in proportion to the required amount for a baseline student success factor.
fundinggap	Calculation of the gap between actual and adequate spending for each school district.
outcomegap	Difference in school district and US National average test scores. The data element will shed light on the school districts with lower levels of academic achievement and will be a good data point to review in comparison to the funding gap.
pov	Census child poverty rate for children 5-17 years of age. The data element will show which school districts have a higher or lower level of poverty and subsequently which have a lower overall availability of academic funds. Here we will also see how areas with high poverty rates are in comparison to others with regards to the funding gap.
State Indicators Database	
inc_effort	A calculation of a state's total educational costs in comparison to their combined state personal income. I could not find this information by school district, but it would also be beneficial to look at by state. What this shows is which states are spending less than their potential. This would be good to look at against the funding gap for all of the districts in a state to see which states were both underfunding their schools and had the resources to be able to allocate more to educational funds. I believe there would be a way to create this same number at the school district level through using a combination of personal income data and location data, plotting each location into a school district.
Common Core Data	
SCHNAM	School name. Each school is tied to a district using a NCES district

	number. Each NCES district name and number would need to be aligned with the districts in the District Cost Database, which also used NCES data as it's starting point.
LEVEL	School level. Delineates the level of education, for example middle or high school.
FRELCH	Number of students eligible to participate in the free lunch act. The data element could be looked at in comparison to the poverty level of the districts for a more robust view of the socio-economic status of the students.
National Assessment of Educational Progress	
Math	Average mathematics scores for elementary, middle, and high schools to be used to gauge scholastic success of an institution.
Reading	Average reading scores for elementary, middle, and high schools to be used to gauge scholastic success of an institution.
Writing	Average writing scores for elementary, middle, and high schools to be used to gauge scholastic success of an institution.
Science	Average science scores for elementary, middle, and high schools to be used to gauge scholastic success of an institution.
Art	Average art scores for elementary, middle, and high schools to be used to gauge scholastic success of an institution.
National Survey of Children's Health	
Flourishing 0-5	Is the young child flourishing? Essentially, is the child in optimal health and meeting or exceeding childhood milestones. The data element would show the number of children in a school district who are maturing mentally and physically at expected levels.
Flourishing 6-17	Is the child flourishing? The data element would show the number of children in a school district who are maturing mentally and physically at expected levels.
ADD/ADHD	Does the child currently have Attention Deficit/Hyperactivity Disorder? The data element would show the number of children in a school district who suffer from these illnesses which can cause difficulties with learning.
Mental or Other	Mental, emotional, developmental, or behavioral problems in children 3-17. The data element would show the number of children in a school district who are struggling with an issue other than ADD/ADHD which could contribute to their scholastic success.
Special Education	Is the child currently receiving services for a special education or early intervention plan? The data element would show the number of students in a school district who are receiving special education services to assist them in reaching a threshold of scholastic success.
School Engagement	How often is the child engaged in school? Does the child care about doing well in school and do required homework? This data point

	would be very interesting to look at in comparison to the school district spending to see if school districts which have higher funding, or higher spending were more likely to have children more engaged in school activities.
Working Poor	Does the child live in a household where the caregivers work full time and have incomes less than 100% of the federal poverty level? The data point would add more insight into the socio-economic status of school districts.
Share Ideas	How well can the child and caregivers speak about things that really matter? This data element would shed some light on the state of the household the child is living in. Open communication between caregiver and child is a healthy environment for a school aged child.
Screen Time	Time the child spends in front of a screen, for example a TV, cell phone, or other electronic device. The data element would assist in indicating the type of home environment a child lives in and help show if they are being set up for success in a schooling environment.
Resilience	Does the child live in a home where the family demonstrates qualities of resilience during difficult times? This data point further shows how the home life of a child could hold sway on their academic success or failure.

Note. Some of the data sources only allowed access to the code books as the actual data files were protected unless licensure was obtained. Through reviewing the code books I was able to gather detailed information on the data elements contained in the data sets.

Analysis

For data analysis I will begin with the basics of classifying all of the data and mapping the data from the differing datasets back to each other. Part of this will entail mapping the addresses of each of the households in the National Survey of Children's Health (NSCH) into their corresponding school districts. As the NSCH does not contain an attribute for school district it will be an involved process to pinpoint the school district each address is slated to attend, and more data resources may need to be leveraged to complete the task. Once all of the data is mapped correctly I will begin looking at data from the district cost database. I would like to

review the school districts with the highest spending per pupil, and those with the lowest spending per pupil. The next item in that dataset in the funding gap. Adequate spending is calculated for each school and the funding gap looks at the difference in the actual spending per pupil and the calculated adequate spending per pupil. This would show which districts are making it a priority to fund their student's education, and which are funding at the bare minimum acceptance level, or are underfunding. Once this is verified at the district level the data will need to be calculated on the state level. Once the data is leveraged by the state the funding gap by state can be compared to the income effort from the State Indicators database. This metric is a comparison of the total education costs by state and the income of said state. Using this we are able to see which states are spending on par with their state incomes, and which are spending less than their potential. Ideally, this metric, next to the statewide funding gap metric, would begin to paint a picture of states, and ultimately school districts, who have a negative funding gap, and are not spending as much on education as their income potential should allow. Once these states are identified further data sources could be leveraged to map state income addresses to the school districts, comparing these school district incomes to district education spending, and them mapping it back to the district funding gaps to pinpoint specifically which school districts in the offending states are key perpetrators.

After gaining a clear view of the funding and spending issues for the states and school districts across the country I would delve further into income issues. I would like to look at the poverty levels for the school districts from the District Cost Database and the number of students signed up for free student lunches from the Common Core Data set. The goal in these data points is to paint a picture of the poverty levels in the various school districts and states. Looking at poverty levels compared to the number of students in the free lunch program would show the

depth of poverty in the district and the concentration of that poverty. Some districts may have similar poverty levels, but one may have a much greater number of students in the free lunch program than the other indicating heavier concentrations of children living in poverty in those districts. This information will be added to the tool bag to continue comparisons on the schools underfunding their students, with higher income potentials, though, I would not expect school districts with high income potentials to also have high poverty rates.

The next step in analysis is to begin looking at student outcomes. Which school districts are performing well in standardized tests both overall, and in specified subjects. To view this data I will use the outcome gap from the District Cost Database and the National Assessment of Educational Progress. The outcome gap shows the difference in a school districts scores for national testing versus the United States averages, and the National Assessment of Educational Progress gives the school averages for math, science, reading, writing, and art. These scores will be looked at against all prior attributes to begin to view if there is a pattern in school test performance, poverty level, funding gaps, spending versus potential, and average spending per pupil. This grouping of data and analysis will make up the package for school district funding and student success. The findings from this aspect of the project can be utilized to benefit public policy, educational budget reforms, and similar areas.

The next area of analysis will be much less straight forward as it is qualitative responses, not a comparison of numbers. “The National Survey of Children’s Health provides rich data on multiple, intersecting aspects of children’s lives including physical and mental health, access to quality health care, and the child’s family, neighborhood, school, and social context” (Learn About the NSCH, n.d.). This is the data that will tie the school funding, poverty levels, and scholastic success into children’s mental health and wellbeing in the school districts. The areas

of the survey I will analyze include whether a child is flourishing, or in optimal physical and cognitive health, is diagnosed as ADD or ADHD, has a cognitive or behavioral disorder not listed, is in special education, and is engaged in school. Looking at these data points I hope to see the level of influence high concentrations of students with disorders potentially preventing them from achieving a high level of scholastic success has on testing grades, and if there are higher concentrations of students with these issues in districts with higher poverty levels. I would also like to look at the special education student levels in comparison to the funding gaps for districts as special education is more costly than the average student. The second set of attributes I would like to review in the National Survey of Children's Health are whether the child lives in a household where the parents are considered working poor, are ideas shared readily among the child's immediate family, how much screen time the child has, and whether the child lives in a family that demonstrates resilience in difficult times. These attributes capture some of the environmental factors which could affect scholastic success for a child. This would again be reviewed against the school district data set.

Upon completion of all of the sub-analyses I hope to have a decent idea of how some of the data correlates and can begin digging into the "why" of the project. Using the data the goal is to see if there are common threads among schools with low scholastic achievement throughout the funding, poverty levels, funding disparities, and child metrics. If these threads can be found the beginnings of a model can begin. The purpose of the model would be to use the historical data from these multitudes of sources to predict how changes could affect student outcomes. For example, in a district with a large income effect, or where the district is not spending as much as it's potential, would closing that gap, based on the child metrics in that district, create a decrease in the gap between said districts testing averages and the national testing averages? Another view

of the project is looking at common threads of child health and wellbeing in areas of highly concentrated poverty and tying those back to the student funding. Could a case be created to aid in promoting public policy to better the lives of the students in the area?

Assumptions and Constraints

The data used will need to be mapped between multiple data sources. Though these are common identifiers in several of the datasets, others do not have this. Without a common identifier the mapping could become messy and there is a possibility for error. The data for the National Survey of Children's Health will not be as complete as the information used for income analysis and poverty levels, or other attributes, so care needs to be taken when comparing and/or contrasting totals from the NSCH to other data sources. It is assumed that the school district data is complete and correct.

Benefits

This analysis could aid in giving those who build public policy insight into the factors influencing educational disparity among different school districts and states. This information could be used to work towards policies to decrease this disparity in the future and the model could help predict how changes to the school system would affect testing averages.

Risks

The riskiest portion of this analysis occurs during the mapping phase where the data from the multitude of data sets is connected. Other risks are bias inherent in the NSCH based on how the survey was conducted, who responded to the survey, and if there is any way to alleviate bias if it is present.

Business Process Changes

Though this project does not affect a particular business it could potentially be used to streamline educational funding in states and change the amounts of funds schools receive in order to improve education in those schools where it is lacking. Currently schools are funded by the local taxes for an area, and if findings show that there is a large difference in educational successes of poverty area schools versus schools without these funding hurdles to overcome it could be used as leverage to better level funding on the student level, not based on the area they grow up in, which is something that is out of their control.

High-Level Timeline and Schedule

The project is very detailed, requires data from multiple sources, and deals in personally identifiable information and the initial stages of preparing the data, mapping it between all sources, and adding location data to the items missing it will take an extended period of time estimated to be 8-12 months. The analysis phase will move quicker at an estimated 3-6 months. The presentation period will follow, but for business processes to change could take upwards of 5-10 years.

Recommendation

Based on the findings above the data is available to get a very robust view of aspects causing educational inequality in United States children. There are many stakeholders who could benefit from this project such as the school boards, local government officials, non-profits, and other firms looking at improving area school systems. Parents and pediatricians could also benefit from insight into this matter. That being said, there will be large gaps in data due to not

every child being reported on the NSCH, the subjectivity of the NSCH, and the inherent bias that come with utilizing that survey as it may be more or less likely for specific groups to have it completed. If the project moves forward it must do so with this understanding. The other aspects of the data are straighter forward and can still provide a clear picture of educational funding.

Proposed Solution

This is such a multi-faceted issue that it is difficult to think of a way to fix it. I would expect to see that the poverty area school districts show less educational acuity than those in higher-income areas. I also expect for poverty level school districts to have more children rating higher in the NSCH scales than higher-income areas. The most interesting aspect I see in this project is the view of which areas are under-spending and which are over-spending, and do those areas have the income capacity to justify the spending. Ultimately a proposed solution would be to divide state income taxes among the schools based on enrollment, or student levels, and not school location. In a perfect world this would equalize education, but as previously stated, so much more plays into a student's academic success that it is hard to know if even that would have a positive effect.

References

Albert Shakner Institute, & Rutgers Graduate School of Education. (2018). District Cost Database. *School Financial Data.org*. Retrieved October 10, 2021, from https://www.schoolfinancedata.org/

Albert Shanker Institute, & Rutgers Graduate School of Education. (2018). State Indicators Database. *School Financial Data.org*. Retrieved October 10, 2021, from https://www.schoolfinancedata.org/

Card, D., & Payne, A. (2000, November 19). School finance reform, the distribution of school spending, and the distribution of student test scores. *Journal of Public Economics*, 35.

Data Resource Center for Child Adolescent Healt. (2019). National Survey for Children's Health. *Child and Adolescent Health Measurement Initiative*. Retrieved October 10, 2021, from https://www.childhealthdata.org/learn-about-the-nsch/nsch-codebooks

Hanson, M. (2021, August 2). *U.S. Public Education Spending Statistics*. Retrieved from educationdata.org: https://educationdata.org/public-education-spending-statistics

Learn About the NSCH. (n.d.). Retrieved from Data Resource Center for Child and Adolescent Health: https://www.childhealthdata.org/learn-about-the-nsch/NSCH

National Center for Education Statistics. (2016). Common Core of Data American Public Schools. *IES NCES*. Retrieved October 10, 2021, from https://nces.ed.gov/ccd/reference_library.asp

National Center for Educational Statistics. (2016). National Assessment of Educational Progress. *The Nation's Report Card*. Retrieved October 10, 2021, from https://www.nationsreportcard.gov/ndecore/xplore/ltt

Paulus, L., Spinath, F., & Hahn, E. (2021, February 23). How do educational inequalities develop? The role of socioeconomic status, cognitive ability, home environment, and self-efficacy along the educational path. *Intelligence*, 11.