

Using Cluster Analysis to Determine Factors of Success In Local Chicago High Schools

Daniel Bulgarelli, Zak Clark, Laura Cuevas, Hannah Niehoff
Loyola University Chicago
April 14, 2020

Abstract:

There have been many methodologies to rank schools in the past. Many of these are based on graduation rates, but they do not show the reason why some schools have greater success than others. The goal of our study is to look at environmental factors that may lead to school success. These factors are critical in finding out relevant solutions so that equality can be achieved in public schooling. We found, by using cluster analysis, several important factors that correlate with school success. These factors give impactful insights into the harsh reality of the inequality in public schooling. By using these clusters of different types of schools, policymakers can make informed decisions when allocating funds and be able to better provide for the school districts and the students within them that are in need of the most help.

1 Introduction:

Chicago Public Schools (CPS) is the third largest school district in the United States. With over 600 schools, including 165 high schools, it is difficult to ensure that all of the schools in the district successfully prepare their students for the future ("About Our Schools, 2018). Schools are grouped by socio-economic status, however, each student, regardless of school attended, must meet the same requirements to advance from one high school grade level to the next.

The purpose of this analysis is to identify the biggest contributing factors to the success of high schools in the Chicago Public Schools system.

There are many different websites and institutions that rank schools based on various factors. While some of those may be accurate, we wanted to take a more in-depth look in order to find out what factors make some schools better than others. Due to the extreme diversity of neighborhoods in Chicago, there is a wide range of public school rankings some public schools are ranked well above others. There are many approaches to ranking schools, however, most school rankings are based on literacy and graduation rates. However, our research intends to discover the impact of environmental factors on schools' success. Neighborhood inputs include factors such as median

household income as well as prevalence of libraries, parks, and crime can have a serious effect on high school students in the area. Additionally, different factors like parental involvement, attendance, and teacher quality can influence a student's experience and therefore their school achievement and scores. Even though the outcomes of school systems in Chicago are well documented, little information is present about what factors actually influence these results. Through this research, we hope to gain a better understanding of what factors determine the success or failure of school systems in Chicago. In this analysis, we hope to bring about actionable information for those in decision making positions as well as the community of Chicago as a whole. It is the purpose of this research to identify areas where certain schools and neighborhoods are lacking and where others are excelling in order to provide valuable insights into the improvement of public education. The information discussed in this paper will have the potential to uncover areas where school systems can progress in order to improve the well-being of the students that attend them.

2 Literature Review:

The Literature surrounding education and student success in the United States is significant. There have been various studies and publications examining key metrics in student success, these span from standardized test scores to qualitative measures like health and wellness. Additionally, school success among different subgroups such as disadvantaged youth and ethnic groups have been analyzed.

In 2016, Crystal Coker published a study with DePaul University, "Neighborhood and School Influences On Academic Achievement and Educational Attainment." The study focused on neighborhood factors when analyzing student success. Unlike previous studies, Coker identified factors that classified neighborhoods as disadvantaged. These factors were not solely based on income level, "Specifically, neighborhood disadvantage is measured as a composite of four census variables: the proportion of families living below poverty, proportion of civilians 16 years and over who are unemployed, proportion of the population 25 years and older without a high school degree, and the proportion of families headed by a single female" (Coker, 2016, p.26). Additionally, Coker used a sample of students from varying neighborhoods that only attended choice schools such as magnet schools, charter schools, open enrollment public schools, and private schools (Coker, 2016, p.19). This sample decreased the significance of school quality and emphasized the importance of individual student success based on their home neighborhood, therefore, examining external factors. There are differing opinions on the advantages placing low-income students in affluent schools. The study found that students from disadvantaged neighborhoods may also be disadvantaged in school and early adulthood, regardless of the type of school attended. While it does not completely explain causation, there are further opportunities for studies and research in this area (Coker, 2016, pp.88-89).

In researching success of the Chicago Public Schools specifically, The University of Chicago published a report in 2017, "The Educational Attainment of Chicago Public Schools Students: 2016." Their metrics heavily emphasize quantitative data such as ACT scores and GPA as well 4-year college enrollment. Their research has found that Chicago Public Schools have seen significant improvement since the beginning of their research in

2016. The University of Chicago has developed a Degree Attainment Index which measures the likelihood that high school freshmen will graduate high school in four years and then graduate college in either six or ten years. These indexes are called the Direct Bachelor's DAI and Bachelor's DAI respectively. The report published in October of 2017, examining Chicago Public Schools, shows improvement in both of these indexes, "We estimate that 16 percent of 2016 ninth-graders, compared to 9 percent of 2006 ninth-graders, will take a direct path to a bachelor's degree by making an immediate transition after high school to enroll in a four-year college and graduate within six years" (Nagaoka, Seeskin, Coca, 2017, p.4). Additionally the Bachelor's DAI for 2016 shows, "The 2016 Bachelor's DAI is 18 percent, 2 percent higher than the Direct Bachelor's DAI. That is, we estimate that 18 percent of 2016 ninth-graders will earn a bachelor's degree by 2026," (Nagaoka, Seeskin, Coca, 2017, p.4). These metrics are essential in measuring the effectiveness of Chicago Public Schools. Additionally, his report further analyzes test scores and metrics in different ethnic groups, the report shows significant disparity with test scores in different ethnic groups. For example, 49% of white males received a 24 or above on the ACT while only 15% of Latino males received scores in that group. Although there are many factors that could explain the difference in scores, it is clear that there is an ethnic disparity (Nagaoka, Seeskin, Coca, 2017, p.17).

Especially, in the city of Chicago, crime is a significant and common barrier with education and school safety. In her study published in October of 2013, Julia Burdick-Will examined Chicago crime data involving school incidents for eight years, from 2002 to 2010. This data examined crime reports and found that, "Of the approximately 100 high schools in Chicago, two thirds called the police to intervene in at least one violent incident on school grounds during the first seven months of the 2009–2010 school year and one quarter of schools called the police more than 17 times during that period." (Burdick-Will, 2013, p.2). Researchers found this data to be significant due to the cognitive stress levels which can hinder cognitive learning. Increased aggression among students and administrators is another effect observed due to higher crime rates at

schools, this can cause a hostile environment and make it increasingly difficult for students to trust peers and feel safe on campus (Burdick-Will, 2013, p.3).

In addition to researching external factors that affect quantitative metrics, some literature focuses on measuring success in using qualitative methods. In 2017, a University of Massachusetts professor published an article with the Harvard Graduate School of Education. Jack Schneider wrote “What Makes a School Great” in which he claims that school rankings miss valuable information. There are many factors that go beyond test scores, but they are much more difficult to measure and compare. He discusses the consequences of heavily relying on metrics like test scores when analyzing school success “Multiple-choice tests communicate nothing about school climate, student engagement, the development of citizenship skills, student social and emotional health, or critical thinking. School quality is multidimensional. And just because a school is strong in one area does not mean that it is equally strong in another. In fact, my research team has found that high standardized test score growth can be correlated with low levels of student engagement” (Schneider, 2017). Schneider further emphasizes looking at factors that “don’t count” in schools such as student satisfaction and wellness (Schneider, 2017).

One of the factors that has steered school rankings and measurements to qualitative measurements is legislation. The National Education Association has several published articles on the impact of educational legislation on school and student success. In December of 2015, President Obama signed the “Every Student Succeeds Act” for education. Since its passing, the National Education Association has analyzed different initiatives being taken nationwide to measure school success. In her article “What Are the Best Measures of School Quality? Educators Speak Out” Brenda Alvarez discusses a poll sent out to 1,200 educators asking them the most important measures of school success. The two most important factors according to the poll were student access to enriching programs including foreign language, fine arts, physical education, learning resources and health and wellness programs (Alvarez, 2016). These kinds of essential school components are constantly overlooked, another

article on the National Education Association’s website, “Accountability After NCLB: Let’s Measure More Important Things Than Test Scores” by Tim Walker, blames the “No Child Left Behind” policies enacted in 2002. The biggest criticism of “No Child Left Behind” is the lack of funding for under-performing schools, “The idea of using a sole metric – test scores – to measure student outcomes was doomed to fail for a host of reasons, not least of all because lawmakers ignored a critical step: providing students in high poverty schools with the necessary supports and resources to help them learn” (Walker, 2015, p.3). The National Education Association goes on to discuss the Opportunity Dashboard that is being enacted to collect data on school district’s enriching programs, student success, and other metrics not related to test scores (Walker 2015, p.4).

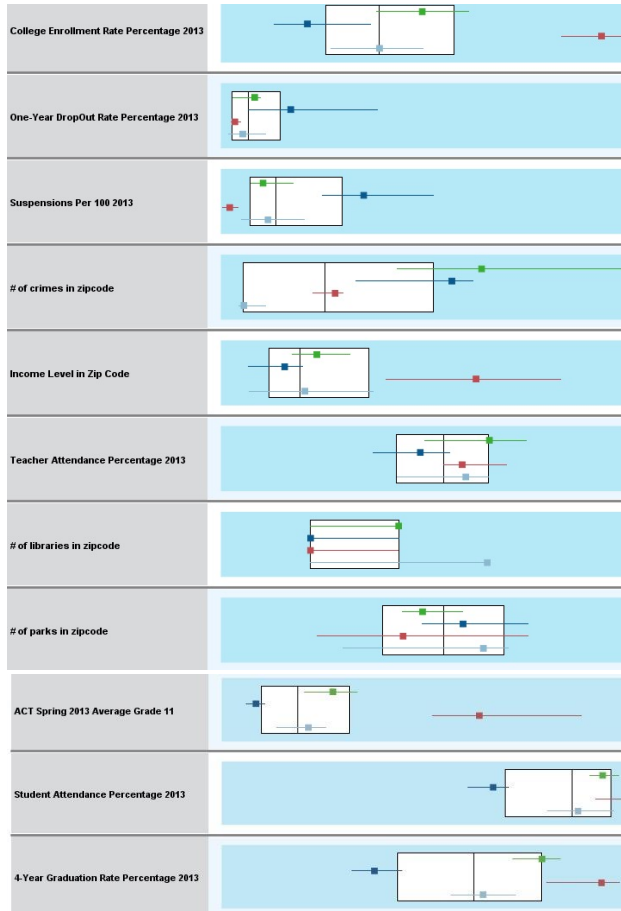
3 Data Description:

In this research, we have compiled 4 individual data sets linked by zip code. The first data set used is a school progress report issued by the Chicago Data Portal in 2012. This data set provides information such as attendance rates, graduation rates, ACT scores, and college enrollment percentages for each individual school district located in the Chicago Area. Out of this data we elected to take the above metrics for Chicago High Schools and use them to determine which high schools were the most successful. In addition to this, a data set also provided by The City of Chicago was used to aggregate libraries within the same zip codes of the school. This was able to provide total amounts of libraries located in close proximity to each of the school districts. Similar to the libraries, The City of Chicago provided a data set showing individual parks separated by zip code. These numbers were aggregated in order to provide a number of parks located within each school district area.

In addition to public services, measures of the social environment in which the schools exist were also needed. The City of Chicago and Social Explorer were used in order to collect information regarding crime and income levels within the zip codes. The City of Chicago crime data included individual criminal activities which were then combined in order to provide crime totals for each area surrounding the High Schools. This allowed us

to gauge levels of activity surrounding the schools and provide insights regarding how this impacts 83 Chicago High Schools spanning across the entire area of the city.

Figure 1: Cluster Input Distribution



4 Models and Results:

In order to analyze the data collected from the City of Chicago, a k-means cluster analysis model was used. A cluster analysis takes the given dataset and groups it based on similarities among the inputs. This was used to group the schools by varying levels of success and allowed major factors that contribute to success to be discovered. 83 schools were included in the dataset used for our cluster analysis. The k-means analysis created five distinct clusters among the Chicago Public Schools. As seen in Figure 2, cluster one contained 12 (14.5%) schools, cluster two contained 2 (2.4%) schools, cluster three contained 26 (31.3%) schools, cluster four contained 28 (33.7%

educational success. Overall complete and accurate data was provided for

Figure 2: Cluster Input Importance

Cluster	cluster-4	cluster-3	cluster-5	cluster-1
Label				
Description				
Size	33.7% (28)	31.3% (26)	18.1% (15)	14.5% (12)
Inputs	ACT Spring 2013 Average Grade 11	ACT Spring 2013 Average Grade 11	ACT Spring 2013 Average Grade 11	ACT Spring 2013 Average Grade 11 ACT Spring 2013 Average Grade 11 Importance = Mean: 14.40
	Student Attendance Percentage 2013	Student Attendance Percentage 2013	Student Attendance Percentage 2013	Student Attendance Percentage 2013
	4-Year Graduation Rate Percentage 2013	4-Year Graduation Rate Percentage 2013	4-Year Graduation Rate Percentage 2013	4-Year Graduation Rate Percentage 2013
	College Enrollment Rate Percentage 2013	College Enrollment Rate Percentage 2013	College Enrollment Rate Percentage 2013	College Enrollment Rate Percentage 2013
	One-Year DropOut Rate Percentage 2013	One-Year DropOut Rate Percentage 2013	One-Year DropOut Rate Percentage 2013	One-Year DropOut Rate Percentage 2013
	Suspensions Per 100 2013	Suspensions Per 100 2013	Suspensions Per 100 2013	Suspensions Per 100 2013
	# of crimes in zipcode	# of crimes in zipcode	# of crimes in zipcode	# of crimes in zipcode
	Income Level in Zip Code	Income Level in Zip Code	Income Level in Zip Code	Income Level in Zip Code

) schools, and cluster five contained 15 (18.1%) schools. Cluster two was excluded from further analysis because it only contained a small fraction of Chicago Public Schools.

Many factors contributed to the grouping of the various schools with the most important being ACT scores, student attendance percentages, and four year graduation rates. Cluster one was determined to be the most successful cluster as it contained schools with the highest test scores, attendance, and graduation rates. Clusters four and five were similar in most aspects. However, based on the most important factors as shown in Figure 2, cluster five is a more successful high school. Cluster three is the least successful highschool based on the variables

included in this analysis. As Figure 1 shows, this cluster contains students from families with the lowest income and has the second highest amount of crime. Cluster three has the lowest teacher and student attendance rates, ACT scores and graduation rates. The levels of importance among the contributing factors as well as the differences between each cluster helped us to determine which factors are correlated with a successful Chicago Public High School.

5 Conclusions and Findings:

Through our cluster analysis, we were able to create distinct groupings between schools that excel and schools that underperform. The first category by which the model drew distinctions from was ACT scores. From best scores to worst scores the clusters ranked were Cluster 1, Cluster 5, Cluster 4, and Cluster 3. Following this we see the same order of clusters in other relevant success metrics such as attendance percentage, graduation percentage, and college enrollment rate. With this same order repeated through all of these measures, it is easy to conclude that the model was able to successfully group together schools that are providing the best educational experience.

A highly significant result the data exhibited was the impact of wealth in a zip code on the quality of a public high school. Schools that are located in zip codes where the residents have a higher average income scored much higher in other significant metrics such as ACT score and college enrollment. Income by zip code was found to be the most significant input as its addition significantly changed the clusters and their characteristics. This conclusion is well supported in previous studies, further reinforcing this correlation.

One surprising metric was the influence of the number of parks in a zip code on the success of a school. Policy makers may believe that more parks in an area will lead to less crime, increased health benefits, and more community involvement. However, our data showed the opposite. Cluster 1, which contains 12 high schools, was found to contain the best and most successful schools. However, the zip codes that these schools are located in contain the smallest number of parks. Cluster 3, which contained 26 high schools, has the second highest number of

parks but contains the high schools that are most lacking in attributes that make a high school excel.

Another surprising result of our analysis was with libraries. A common misconception is that more libraries would result in better grades and better results for the students. By analyzing the number of libraries in a school's zip code, we did not find any correlation between libraries and academic success. This is certainly counterintuitive, but it shows that policy makers should allocate money to different areas to improve grades and high school success. Of course, this assumes that this is the goal of policy makers when funding libraries.

The results that we uncovered are vital to policy makers, who allocate funds to communities. Although further research would be necessary in order to establish causation with parks and libraries, our results show important and unexpected findings. Policy makers' focus on these findings and further investigations could help improve education quality metrics in certain zip codes. This will make for a more equitable and just society for all of the Chicagoland area. This method of analysis may also be useful to policy makers and researchers in other cities across the United States. By analyzing similar datasets on public high schools in their cities, they could also find significant results to further improve education metrics.

6 Recommendations:

In the future, researchers could analyze the determining factors of success for high schools on a larger scale. This research used data on Chicago Public High Schools, and we do not believe that our results can be extrapolated to the rest of the United States. Chicago is a fairly large city, with unique characteristics. While it may be difficult to determine and analyze similar factors across countries, it could be done across the United States. Future research should consist of data from thousands of high schools in various regions of the country. Another recommendation for future researchers is to gather more complete and comprehensive data on the high schools in Chicago. Data is essential to gathering actionable information and insights in order to improve school quality. The city of Chicago has public data on Chicago public high school achievement scores including graduation rates,

attendance rates, and ACT scores etc. However, the data exhibited large data quality issues in that various schools did not have data available for a great deal of categories. The schools that did not have complete data were not included in our study. This makes it impossible to fully evaluate school quality based on the dataset. The lack of information for these schools is a prohibitive factor for in-depth analysis. We strongly recommend further investigation by the city of Chicago in order to find this crucial information.

Another way to analyze Chicago's high schools in a more in depth manner is to include zip codes that the schools receive students from. Our analysis only looked at the zip codes that the schools are located in, but there may be students from outside of these zip codes. This could potentially change the results of our study.

References:

“About Our Schools,” “Stats” (2018) Chicago Public Schools. <https://cps.edu/About_CPS/At-a-glance/Pages/Stats_and_facts.aspx>

Alvarez, B. (2016) *What Are the Best Measures of School Quality? Educators Speak Out*. National Education Association.

Burdick-Will, J. (2013) *School Violent Crime and Academic Achievement in Chicago*. *Sociology of Education*, 86(4), pp.343–361.

Chicago Park District. (2016) *Parks - Chicago Park District Park Boundaries (current)*. Electronic dataset. City of Chicago. <<https://data.cityofchicago.org/Parks-Recreation/Parks-Chicago-Park-District-Park-Boundaries-curren/ej32-qgdr>>

Chicago Public Library. (2019) *Libraries - Locations, Hours and Contact Information*. Electronic dataset. City of Chicago. <<https://data.cityofchicago.org/Education/Libraries-Locations-Hours-and-Contact-Information/x8fc-8rcq>>

Chicago Public Schools. (2013) *Chicago Public Schools - High School Progress Report (2013-2014)*. Electronic dataset. City of Chicago.

<<https://data.cityofchicago.org/Education/Chicago-Public-Schools-High-School-Progress-Report/2m8w-izji>>

Coker, C. (2016) *Neighborhood and School Influences On Academic Achievement and Educational Attainment*. Chicago: College of Science and Health at DePaul University.

Median Household Income in 2013 Inflation Dollars. (2013) Electronic dataset. Social Explorer. <<https://www.socialexplorer.com/a9676d974c/explorer>>

Nakaoka, J., Seeskin, A. & Coca, V. (2017) *The Educational Attainment of Chicago Public Schools Students: 2016*. Chicago. The University of Chicago.

Police. (2020) *Crimes - 2020*. Electronic dataset. City of Chicago. <<https://data.cityofchicago.org/Public-Safety/Crimes-2020/qzdf-xmn8>>

Schneider, J. (2017), *What Makes a Great School?*. Harvard Graduate School of Education.

Walker, T. (2015) *Accountability After NCLB: Let's Measure More Important Things Than Test Scores*. National Education Association.