Comparable School Districts – School Report Card Characteristics

For this agenda topic, the Board will be given the opportunity to discuss and identify those characteristics deemed necessary and appropriate to determine which Illinois school districts could be considerable "comparable" to District 90 (with the goal of comparing student academic performance).

District statistician, Phil Earvolino, will join the conversation. He has been asked to provide explanation for a recommended statistical approach that will both yield accurate outcomes and be relatable for stakeholders across the school community. At the meeting, we will share some suggested school district characteristics that could be used when conducting a comparative analysis and ask that Board members come to general consensus about them.

Please find attached several exhibits that may be useful to frame your thinking. The first attachment includes the list of possible characteristics for Board consideration, as referenced above.

Additional exhibits include:

- Index from the *ISBE Public Business Rules 2023 Report Card Metrics* document
- Listing of characteristics used to discern comparable districts, sourced from the Ohio State Department of Education
- Nebraska Department of Education *Methodology to Compare Districts and Schools: A Technical Report* (January 2019)
- LA Times article: "Similar School Rankings are a Boost for Some Districts" (January 27, 2000).

Once the Board has come to general consensus about the essential characteristics and approved the statistical approach, Phil will conduct analyses and compile the results. The findings from the investigation will be provided to the Board and community later this spring.

Please feel free to contact me directly if you have any questions, comments, or concerns in advance of the conversation.

District 90

Potential Initial Variables for Consideration of Comparable Schools

Student enrollment
Instructional expenditures per student
Percent White students
Percent Black students
Percent Hispanic students
Percent Asian students
EAV per pupil
Percentage of students in top two IAR performance tiers (Meets/Exceeds)
Percent low income
Other

Illinois
State Board of
Education

Public Business Rules
2023 Report Card Metrics

Contents

SOURCE: OHIO DEPARTMENT OF EDUCATION

Data Sources for 2023 Similar Districts

Data Element	Year	Source	Comments
Enrollment (FTE)	School Year 2021-2022	Ohio Department of Education, EMIS	Full-time equivalent (FTE) enrollment of district
% Population Living in Urban Areas	2010	U.S. Census Bureau, Decennial Census	Residing in urbanized areas or clusters as defined by Census Bureau
Population Density per Square Mile	2017-2021	U.S. Census Bureau, American Community Survey 5-Year Estimates	Total population divided by land square miles (from Census TIGER file)
Median Adjusted Gross Income ($)	Fiscal Year 2021	Ohio Department of Taxation, Tax Data Series	Based on state tax returns
% of Economically Disadvantaged	School Year 2021-2022	Ohio Department of Education, EMIS	FTE of students with economic disadv. flag as percentage of total FTE
% Adults with College Degree	2017-2021	U.S. Census Bureau, American Community Survey 5-Year Estimates	Pop. age 25-64 with a bachelor's degree or higher
% Racial Minority Enrollment	School Year 2021-2022	Ohio Department of Education, EMIS	FTE of "minority" students (i.e., non-white or Hispanic) as percentage of total FTE

Methodology to Compare Districts and Schools: A Technical Report

January 18, 2019

Prepared by

Office of Data, Research and Evaluation
Nebraska Department of Education

Table of Contents

Note of Caution

The methodology described in this document represents *one* approach to constructing a group of similar peers for each school and school district in Nebraska. Other methods could also be used. As such, we caution readers to interpret the similar peer information with care. When evaluating school and school district data, persons should consider a mix of reference points as a means of triangulation. Other reference points might include, for example: the state average, statistics for those schools and school districts geographically closest, statistics for schools receiving similar supports and services, and those with the most similar membership counts.

Limitations

Developing similar peer groupings is designed to enable users to conduct more thoughtful comparative analysis. Despite the benefits to this approach, there are limitations to the use of any grouping methodology. Specific limitations to the approach employed here include:

- The similar peer calculation does not include a measure of geographic distance (although users can select geographic distance as a separate parameter using the NEP compare feature). Many schools and school districts tend to compare themselves with surrounding schools and school districts. The similar peer method does not necessarily include geographically close districts in the comparison grouping because neighboring districts might not truly be the "most similar" districts in the state. On the other hand, some variables included in the similar peer calculation tend to reflect regional conditions.
- The similar peer method deliberately selects only the 12 schools or school districts "most similar" as the standard for comparison. However, some schools and districts are more "unique" than others. In some cases, "similarity" to other schools or school districts – even among peers – can be large.
- It is also true that some schools or school districts tend to look like many other schools or school districts, so the cutoff of 12 captures those schools or school districts that are extremely similar according to the chosen dimensions. Still, schools or school districts can closely resemble many other schools or school districts beyond the cutoff of 12.

Acknowledgments

- Dr. Ashok Samal, Professor, Department of Computer Science and Engineering, University of Nebraska-Lincoln (UNL) for providing invaluable guidance on variable identification, data management, and clustering approaches.
- David Drozd, Research Coordinator, Center for Public Affairs Research (CPAR), University of Nebraska at Omaha (UNO) for providing expert insight into Census data and procuring the desired Census data for this project.
- Office of Policy and Research, Ohio Department of Education, for sharing their experience and method of determining similar districts in their state.
- Subject Matter Experts at the Nebraska Department of Education (NDE) for sharing their knowledge on the appropriateness of specific data elements for this project.

Keywords: *Similar Districts; Similar Schools; Nebraska Education Profile; Census Data; Euclidean Distance; Geographic (Haversine) Distance*

Introduction

The Nebraska Education Profile (NEP) website has been undergoing major enhancements, and thus the need to identify and compare similar peer districts and schools. This would provide utility for any given district or school as they evaluate their performance relative to that of the entire state, and relative to that of other districts or schools that are similar to them on a variety of measures – peers. Additionally, groups of districts or schools that are geographically close to each other are also determined to allow for comparisons between districts or schools within the same geographical area. This technical report details the methodology behind these similar peers and geographic groupings.

Similar Peer Districts and Schools

Design and Methods

In order to operationalize "similarity," a combination of variables that uniquely describes each district or school was identified. These variables were selected due to their relevance, availability, and persistence. Table 1 describes the list of 27 variables that were selected to describe any given district or school.

Table 1. Variables used to compare similarity between districts and schools.

Variable	Description	Source
Membership	Number of students enrolled	NDE
Attendance Rate	Average student attendance rate	NDE
Graduation Rate	4-year graduation rate for the 2016-2017 cohort	NDE
FRL Rate	Percentage of free-and-reduced lunch students	NDE
Minority Rate	Percentage of non-White students	NDE
Homeless Rate	Percentage of homeless students	NDE
LEP Rate	Percentage of English language learners	NDE
Migrant Rate	Percentage of migrant students	NDE
ELA Percent Proficient	Percentage of students proficient in ELA	NDE
Math Percent Proficient	Percentage of students proficient in Math	NDE
Science Percent Proficient	Percentage of students proficient in Science	NDE
Teachers With Masters Percent	Percentage of teachers with at least a Master's degree	NDE
Average Years Teaching Experience	Average number of years taught by teachers	NDE
Unduplicated Suspensions	Number of students with suspensions	NDE
Unduplicated Expulsions	Number of students with expulsions	NDE
Land Valuation	Annual land valuation sent out from the County Treasurer's office of the district	NDE

Variable	Description	Source
Per Pupil Cost by Average Daily Membership	Total annual costs divided by the average daily membership for the district	NDE
Grand Total of All Receipts	Amount of all receipts/revenue received by the district in a school year	NDE
Median Household Income	Median household income in the past 12 months (in 2016 inflation-adjusted dollars)	Census-ACS 2012-2016
Per Capita Income	Per capita income in the past 12 months (in 2016 inflation-adjusted dollars)	Census-ACS 2012-2016
Gini Index	Gini index of income inequality	Census-ACS 2012-2016
Percent Age 25+ With Bachelor's Degree or More	Percent of population 25 years and over with at least a Bachelor's degree	Census-ACS 2012-2016
Labor Force Participation Rate	Percent of population 16 years and over in the labor force	Census-ACS 2012-2016
Unemployment Rate	Percent of population 16 years and over who are unemployed	Census-ACS 2012-2016
Total Population	Population in the district	Census 2010
Land Area	Area in square miles	Census 2010
Population Density	Density per square mile of land area	Census 2010

In creating the district and school data sets from various data sources, a number of challenges surfaced. First, the latest data from NDE was the 2016-2017 school year, while the latest data from the Census was from 2010, and from 2012-2016. Although the Census data lagged behind NDE's data on the districts and schools, the Census data was still used since the variables described community characteristics (e.g., median household income, land area, etc.) that would likely not have changed as frequently as the school characteristics (e.g., membership, attendance rate, etc.).

Second, the Census data was only collected at the district-level, and not at the school-level. However, since the community characteristics of a given district would reflect that of the schools within the district, the same Census data was used at the school-level. This implied that all schools within the same district would, for example, have the same unemployment rate as that of the district. Three pieces of finance data were also collected at the district-level only by NDE: land valuation, per pupil cost by average daily membership, and grand total of all receipts. By the same logic aforementioned, district-level information was used for the schools within the same district.

Third, there were a number of districts that were consolidated after the Census data was collected. In these cases, the originating districts were first identified in the Census data, and the average values of the Census variables were then calculated to inform the Census variables for the new consolidated district.

Once the aforementioned decisions were made, a data split was performed on only the school data file. The school data file was split into three separate data files to reflect the differences among elementary, middle, and high schools. The number of students with expulsions was found to have very little variability across the schools (due to many zero values) and was thus removed from all school data files. Only one variable was not available to describe the elementary and middle schools, namely, graduation rate which was only applicable to high school students. With three school data files, and one district data file, the analyses to identify similar districts and schools commenced.

Analytic Approach

Each district or school was compared to every other district or school by using a distance measure between each pair of districts or schools. This Euclidean distance measure was calculated as a summary index using the formula shown below:

$$d_{euc}(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

In the formula above, d represents the distance between any two districts or two schools x and y on each variable i (i.e., every variable shown in Table 1). Due to the wide differences in the ranges of values across the variables, each variable was scaled prior to computing the Euclidean distance.

Thus, for each district or school, the districts or schools with the shortest distances to it are grouped together. This is because the shorter the Euclidean distance between two districts or two schools, the more similar they are.

Geographical Area

Design and Methods

The addresses for each district and school building were first converted into latitude and longitude information. Once this was done, the geographic distance between every pair of districts and every pair of schools was calculated using the Haversine distance measure. Note that the school data file was split into three separate data files to ensure that similar school types were being compared to each other. For example, elementary schools were only compared with other elementary schools in terms of geographic distance. The same held true for middle schools and high schools as well.

Table 2. Variables used to describe geographic location for districts and schools.

Variable	Description	Source
Latitude	North-South geographic coordinate	Google Maps
Longitude	East-West geographic coordinate	Google Maps

Analytic Approach

Each district or school was compared to every other district or school by using a geographic distance measure between each pair of districts or schools. This Haversine distance represents the distance between two coordinates on a sphere and was calculated using the formula shown below:

$$d_{hav}(x,y) = 2r\ \sin^{-1}\left(\sqrt{\sin^2\left(\frac{\varphi_y - \varphi_x}{2}\right) + \cos(\varphi_x)\cos(\varphi_y)\sin^2\left(\frac{\lambda_y - \lambda_x}{2}\right)}\right)$$

In the formula above, *d* represents the geographic distance between any two districts or two schools *x* and *y*, with φ representing the latitude and λ representing the longitude.

Results

The results of this work can be found as an interactive display in the Nebraska Education Profile website: http://nep.education.ne.gov/. Once a district or school is selected from the dropdown menu on the main page, the "Compare" feature can then be selected to show 10 other districts or schools that are most similar or geographically closest to the referent district or school. For questions or comments regarding the use of this feature, please reach out to NDE.Research@nebraska.gov.

Contributors

This research effort was conducted by the following researchers at the Office of Data, Research and Evaluation at the Nebraska Department of Education:

- Matt Hastings, Ph.D., Senior Administrator
- Hongwook Suh, Ph.D., Psychometrician Lead
- Justine Yeo, Statistical Research Analyst
- Kunal Dash, Statistical Research Analyst
- Fisayo Adeniyan, Research Assistant

Appendix

All distance calculations were computed using R, a statistical software. The syntax is shown in the tables below. While only the syntax for the district data is presented, the same syntax was also applied to all school data files.

Table 3. Syntax for calculating Euclidean distances for every pair of district.

```
###Euclidean Distance
###District Data

#install.packages("ggplot2")
library(ggplot2)
#install.packages("factoextra")
library(factoextra)
#install.packages("xlsx")
library(xlsx)


getwd()
setwd("District Data")
getwd()
district <- read.csv("District Data v0.09.csv")
head(district)
#district <- na.omit(district)

district[,-c(1)] <- scale(district[, -c(1)])
head(district)

districtdistance <- dist(district, method="euclidean")
as.matrix(districtdistance)
as.matrix(districtdistance)[1:6, 1:6]
distanceframe <- round(as.matrix(districtdistance), 5)
str(distanceframe)

fviz_dist(districtdistance)

write.csv(distanceframe, "District Euclidean Distance.csv")
```

Table 4. Syntax for converting addresses to latitude and longitude coordinates, and for calculating Haversine distances for every pair of district.

```
###Geocoding
###District Addresses Data

#Install necessary packages
#install.packages("tidyverse")
```

```
library(tidyverse)
#install.packages("ggmap")
library(ggmap)
#install.packages("geosphere")
library(geosphere)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("xlsx")
library(xlsx)

#Set working directory
getwd()
setwd("Geographic Distance")
getwd()

#Import data with addresses
adddistrict <- read.csv("District Address v0.01.csv", stringsAsFactors = FALSE)
head(adddistrict)
adddistrict <- na.omit(adddistrict)

#Convert addresses to longitude and latitude
?mutate_geocode
geodistrict <- mutate_geocode(adddistrict, Location)
head(geodistrict)

#Check status of query counts from Google Maps (limited to 2500 queries per day)
geocodeQueryCheck()

#Export data with longitude and latitude columns appended
write.csv(geodistrict, "District Geocode v0.01.csv")

#Import data with longitude and latitude columns only
district <- read.csv("District Geocode for Distances v0.01.csv")
head(district)

#Drop agency name which is the first column in the data
district2 <- district[,-c(1)]
head(district2)

#Calculate distance between every pair
distance <- distm(district2, fun=distHaversine)

#Convert distances into a matrix
as.matrix(distance)
```

```
as.matrix(distance)[1:6, 1:6]
str(distance)

#Export matrix of distances
write.csv(distance, "District Geographic Distance v0.01.csv")
```

Los Angeles Times

CALIFORNIA

Similar-School Rankings Are a Boost for Some Districts

BY DARYL KELLEY

JAN. 27, 2000 12 AM PT

TIMES STAFF WRITER

Ten of 11 schools in the Hueneme Elementary District rated among California's underachieving campuses when the state released its first set of rankings Tuesday. So why was Supt. Robert Fraisse smiling Wednesday morning?

The reason is that no Ventura County district did better than Hueneme when compared with communities of similar income, education and English-language proficiency.

Credit good, well-paid teachers, Fraisse said, and extensive after-school instruction in English for the predominantly Latino district, where nearly half the students speak halting English.

Filled with uniformed students, Hueneme schools--located in south Oxnard and Port Hueneme--are also marked by strict discipline and extensive use of computers, he said.

Four Hueneme campuses received the state's top ranking of 10 when compared with

similar schools. Four more had similar-school rankings of 8 or 9 on a 1-to-10 scale.

"I'm just delighted they're looking at these rankings from two different angles," Fraisse said. "I'm extremely proud of our ranking relative to similar schools."

Districts throughout the county were reacting Wednesday not only to their overall state rankings but also to the separate ranking that attempts to reflect how districts--rich and poor--compare with their own kind.

And some administrators were not happy.

In Thousand Oaks, where 14 of 26 schools received the state's top ranking of 10, officials had to explain why their high-performing campuses did so poorly against schools in other well-educated, affluent communities. They concluded that the similar-school rankings were bogus because data the state used to make the comparisons were incomplete.

"Those results are sort of spurious," Conejo Valley Supt. Jerry Gross said. "But the state is now saying they're going to rerun those rankings with complete data and revise our rankings."

The problem in Conejo and several other districts was that educators did not provide information on two key variables used to determine the similarity of schools--student movement in and out of school and the number of students whose families are poor enough to receive subsidized lunches.

That background information was listed as optional on last year's Stanford 9 basic-skills test, the cornerstone for the first-time rankings of California schools, so many districts did not take the time and considerable effort to provide it, Gross said.

State education officials said Wednesday that they will require complete data the second time around, and that will be reflected in reports next fall.

They said they had received dozens of complaints--most from affluent districts--about the similar-school rankings and are considering rerunning them this spring once complaining schools provide the extra information.

But Bill Padia, director of policy and evaluation for the state Department of Education, said he believes that the similar-school rankings are already a good indicator of relative performance because so many variables were used. Even if one or two measures were off, the others would tend to balance out the equation, he said.

"Here's the deal--the data are not perfect because it's the first time we've ever done anything like this," Padia said. "But when you add in eight variables and you get three or four back, then you have a pretty good idea of where these schools are and which ones you can compare.

"Schools that always get these high rankings are not used to being compared with

other schools like them that are out of the area," he added. "So suddenly the competition heats up, and it's uncomfortable. It's like when you graduate and go to Stanford or Berkeley and suddenly you're not the smartest in the class anymore. So we're getting a fairly disturbed reaction from the high-level schools."

*

Among the eight variables used to determine the similarity of schools are pupils' mobility, ethnicity, socioeconomic status and whether they speak limited English. The other variables are class size, and the percentage of teachers who have full credentials and those with only emergency credentials.

Responding to Conejo's specific complaints, Padia said the number of students who move into and out of a school--so-called student mobility--is one of the least important variables if data are otherwise complete.

Determining how many students receive subsidized lunches can be more crucial, he said, because it is one of only two variables used to determine a student's economic status. The second is parents' education level. If one of the two is provided, he said, the second is not necessary to fill out a school's profile.

But Gross wasn't buying Padia's explanation.

"I'm anxious to see how Conejo really ranks relative to similar schools," he said.

Around the county, other administrators were left either scratching their heads about similar-school rankings or celebrating them.

At the Oxnard High School District, it was mostly celebration.

The district's four high schools where income and education levels are lowest--Oxnard, Channel Islands, Rio Mesa and Hueneme--scored well compared with similar schools.

Oxnard High, which scored only average on the overall state rankings, was the county's only high school to receive a 10 in similar-school rankings. Conversely, the district's fifth high school, Camarillo, ranked a high 9 overall, but only a 5 when compared with similar affluent schools.

"We provided all the background information for all five schools," Assistant Supt. Gary Davis said. "And this just shows that compared to like-kind schools, we performed very well. But we certainly want to engage in conversation with Camarillo. I'm sure the staff there is as concerned as the district staff."

*

Across town, Oxnard elementary's Rose Avenue School, partially a magnet for gifted students, also stood out. It scored a 10 compared with similar schools, but only a 5 compared with all state schools.

In the nearby Ocean View Elementary District, where about half the students speak limited English, three of four schools rated only average against all schools. But they received 9s or 10s compared with comparable schools, even though the district did not report low parent education levels or mobility rates.

"We're not quibbling with these rankings," said Jeff Chancer, associate superintendent. "When kids are required to take this test and they can't read English, how are they going to compete with these affluent districts?"

But they did compete with peers, he said, because of extraordinary teaching.

In upscale Camarillo, Pleasant Valley elementary district officials were celebrating their high overall rankings, but also looking hard at why some high-ranking schools ranked low compared with similar campuses.

"We're still trying to get a handle on it," said Barbara Wagner, director of instructional programs. "We talked about it today with the principals. We're looking at the background data from each school to see if it is clean."

*

So far, she said, she's discovered that parent education data are not as good as it should be, since that information is voluntary and is returned at different rates at different schools.

"But we're not looking for any excuses," Wagner said. "We want to work with this information."

Officials in the white-collar suburban enclave of Oak Park, where all five schools exceeded the state target for high performance, questioned the reliability of the data to determine its ranking against similar schools.

That was partly because of the rankings of Oak Hills Elementary, which ranked a 10 for performance compared with all California schools, but only a 1 when compared with schools like itself. The other two district elementary schools scored 10s overall, but just 5 against similar schools.

District curriculum consultant Sharon Morgan said the results raise serious questions

about the validity of information used to decide what schools are similar. All three elementary schools are alike in demographics and all scored high on basic skills tests, but had divergent similar-school rankings, she said.

"It just begs questions," she said. "I've had two calls today [from parents] saying the same thing. They were pleased with our performance, but curious about the difference in the two sets of rankings."

*

One problem Morgan said she found was that the state shows no Oak Park children with subsidized lunches, and that is just wrong.

"I don't know enough yet to say if it was [the school's] error or the state's error in not picking it up," she said.

Regardless, Morgan said the district is pleased with its top statewide ranks compared with all schools.

"So the [similar-school] rankings really become a curiosity more than a concern," she said. "The bottom line is that it's a number, not a thorough analysis of the quality of the program."

*

* EDUCATION GAP

The gulf between the best and worst is the same as between rich and poor. A3