Comparison Schools

The Oregon Department of Education (ODE) uses an iterative process to identify comparison schools. These are institutions from the same type (i.e., elementary, middle, high, and combined[1] schools) with similar student demographics and student enrollment sizes. The iterative process consists of the following parts: inclusion rules, student demographic variables, component calculation, Euclidean distance, size filter, and the selection of the twenty most similar schools.

Inclusion Rules

The ODE includes the following institutions: (a) schools that receive a report card, (b) schools with student enrollments greater than or equal to 40 students (as of the first school day in May), and (c) schools where the highest grade offered is at least the 4th grade.

Student Demographic Variables

The identification of comparison schools relies on the following student demographic variables:

- The percent of students identified as having access to free or reduced price meals[2]
- The percent of students identified as ever English learners[3]
- The percent of students identified as belonging to an underserved racial/ethnic group[4]
- The percent of students identified as mobile within the school year[5]

Component Calculation

The ODE derives two components from the four student demographic variables using principal components analysis (PCA). PCA is a multivariate statistical technique that employs a weighted linear

[1] Combined schools are schools that are a combination of high school grades and any grades 7 and lower.

[2] These are students who are eligible to participate in the free/reduced price meal program or participate in the free/reduced price lunch program at any time during the school year. This also includes students who are enrolled in a school that offers meals at no charge to all students. Note that this measure is formerly known as economically disadvantaged.

[3] These are students who participated (and were reclassified) or are currently participating in a program to acquire academic English.

[4] These are students who are American Indian/Alaska Native, Black/African American, Hispanic/Latino, or Native Hawaiian/Pacific Islander.

[5] These are students who experience one or more of the following: (a) attends more than one Oregon public school during the school year, (b) enters the Oregon public education system late (i.e., after the first school day in October), (c) exits the Oregon public education system early (i.e., on or before the first school day in May without earning a diploma, certificate, etc.), and (d) has significant gaps in enrollment during the school year totaling ten or more consecutive week days.

combination of a number of similar variables (e.g., student demographic variables) to produce a smaller set of uncorrelated and independent components. The aim of PCA is to reduce the amount of data by retaining the components[6] that explain the most variation[7] in the original variables.

Table 1. Demographic Variable Weights for the Two School Components

Demographic Variable	Component 1	Component 2
Access to Free/Reduced Price Meals	.223	.360
Ever English Learner	.454	-.187
Underserved Race/Ethnicity	.437	-.056
Mobility	-.142	.851

As part of the PCA, each institution receives two component scores. These scores are the sum of the products between the student demographic variables and the weights. The ODE uses the two component scores for each institution to calculate the Euclidean distance between schools.

Euclidean Distance

Euclidean distance is the distance between two points. The ODE weights each institution's component scores[8] and creates a distance matrix to calculate the Euclidean distance for all pairs of institutions within each type. For example, the calculation of Euclidean distance between School A and School B is $\sqrt{(C1_A - C1_B)^2 + (C2_A - C2_B)^2}$ where C1 is the weighted first component score and C2 is the weighted second component score. As an illustration, let's suppose that School A has $C1_A = 1$ and $C2_A = 1$, and School B has $C1_B = 3$ and $C2_B = 3$. The Euclidean distance between School A and School B would be $\sqrt{(1-3)^2 + (1-3)^2}$ or 2.828. The scatterplot below shows School A, School B, and their Euclidean distance.

[6] The ODE retains components where the Eigen values were greater than or equal to one (i.e., this is Kaiser's criterion).

[7] The PCA model in 2017-18 retained two components which explained 86 percent of the total variation.

[8] The ODE weights each component by its Eigen value. The school Eigen values for the first and second components were 2.363 and 1.066.

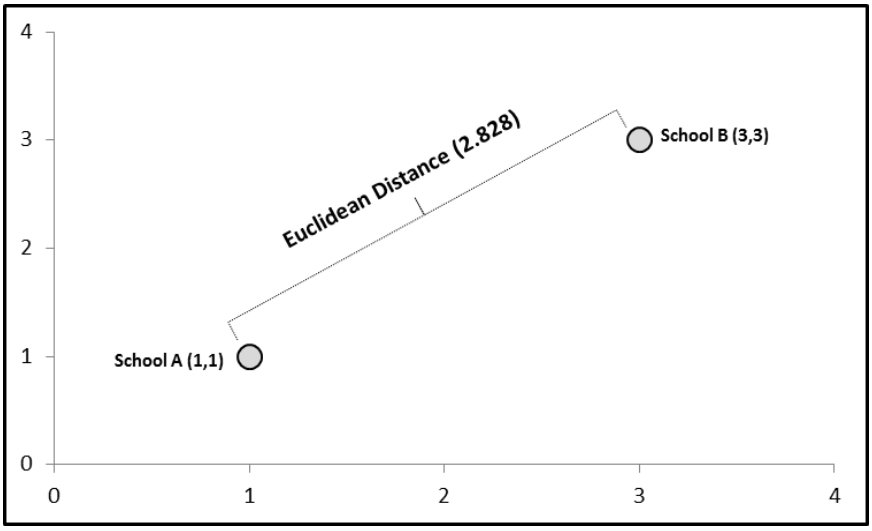

Figure 1. Euclidean Distance between School A and School B

After calculating the Euclidean distance between all pairs of institutions, the ODE removes all comparators where the Euclidean distance between the focal and comparison institutions is greater than 1.75. This considerably reduces the number of comparators to those with the closest proximity (regardless of student enrollment size). The scatterplot below shows an illustration of School B's comparison schools after applying the maximum Euclidean distance rule.

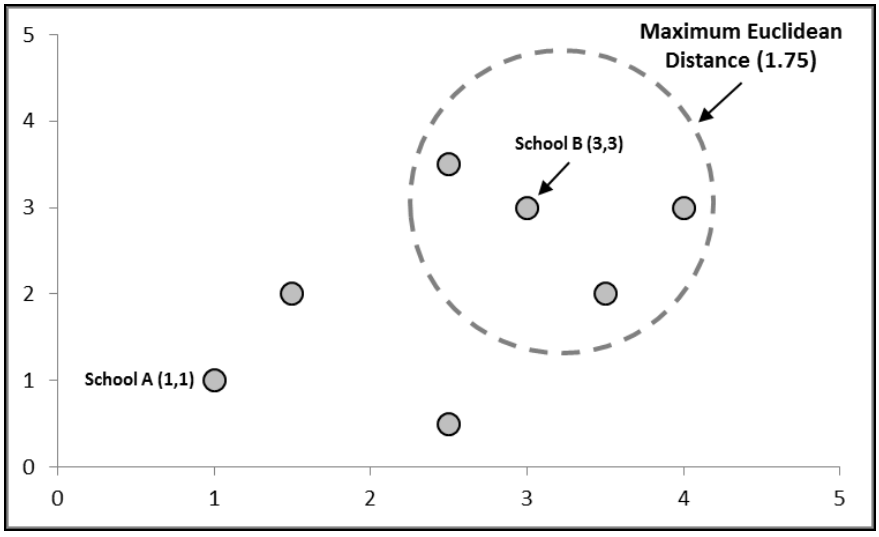

Figure 2. Maximum Euclidean Distance for School B

Size Filter

The ODE applies a size filter to remove any comparators with considerable student enrollment size differences (i.e., sizes either substantially larger or smaller than the focal school). The size filter removes comparators where the size difference in terms of student enrollment between the focal and comparison institution is less than 50 percent of the student enrollment size of the focal school and the comparison school. The following table shows an example of the size filter using School B and its six comparison schools. Note that the size filter removed two schools (i.e., School E and School G) because their student enrollment size was either substantially smaller or larger than School B.

Table 2. Application of Size Filter to School B and Corresponding Comparators

Focal School Name	Comparison School Name	Euclidean Distance	Focal School Size	Comparison School Size	Size Difference as a % of Focal School	Size Difference as a % of Comparison School	Remove (Yes or No)
School B	School C	0.23	500	513	2.6	2.5	No
School B	School D	0.37	500	681	36.2	26.6	No
School B	School E	0.46	500	150	70.0	233.3	Yes
School B	School F	0.65	500	567	13.4	11.8	No
School B	School G	0.72	500	1,500	200.0	66.7	Yes
School B	School H	0.84	500	892	78.4	43.9	No

Twenty Most Similar Institutions

Removing comparison schools using Euclidean distance and the size filter greatly decreases the number of comparators; however, while this left many schools with less than 20 comparators (and a handful with none[9]), a large number of institutions (especially elementary schools) continue to have a large number of comparators. Thus, for focal schools with greater than twenty comparators, the ODE selects the twenty comparison institutions with the greatest proximity to the focal school as measured by the Euclidean distance. These final comparison institutions (twenty or less) are the comparison schools with the most similarity to the focal institutions according to student demographics and enrollment size.

For More Information

Contact: W. Joshua Rew at josh.rew@state.or.us

[9] These schools are very unique and do not have comparators with reasonable Euclidean distance.