Calculation of Comparison Schools for the 2014-15 School Report Cards

The school report cards display the like-school averages for several indicators including English language arts and mathematics achievement, 9th grade on track, dropout, four-year cohort graduation rate, five-year completer rate, and college enrollment among many others. The intention behind the display of the like-school averages is to allow for the comparison between a focal school's performance and the performance of schools from the same school type (i.e., elementary, middle, high and combined[1] schools) with similar student demographics and student enrollment sizes. The Oregon Department of Education (ODE) used an iterative process to identify a focal school's group of comparison schools. This 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 included the following schools: (a) schools that receive a report card, (b) schools with student enrollments greater than or equal to 40 students (as of the 1st 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 relied on the following student demographic variables:

- The percent of students identified as economically disadvantaged.
- The percent of students identified as ever English learners[2].
- The percent of students identified as belonging to an underserved racial/ethnic group[3].
- The percent of students identified as mobile within the school year[4].

Component Calculation

The ODE derived two components from the four student demographic variables using principal components analysis (PCA). PCA is a multivariate statistical technique that employs a weighted linear 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 retain the components that explain the most variation in the original variables.

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

[2] These are students who were ever eligible for or participating in a program to acquire academic English.

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

[4] These are students who experienced one or more of the following: (a) attended more than one Oregon public school during the school year, (b) entered the Oregon public education system late (i.e., after October 1), (c) exited the Oregon public education system early (i.e., on or before May 1 without earning a diploma, certificate, etc.), and (d) had significant gaps in enrollment during the school year totaling ten or more consecutive week days.

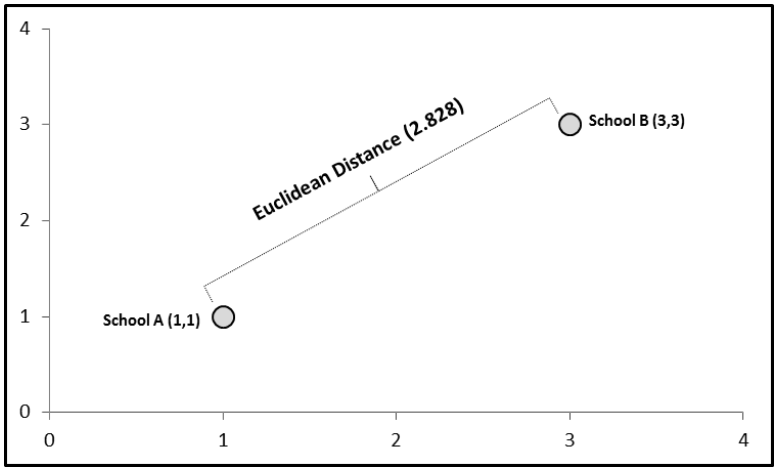
The ODE transformed each student demographic variable into a z-score and used PCA to retain the components[5] that explained the most variation in the student demographic variables. The PCA model retained two components which explained 86 percent of the total variation. The weights for the two components were the following:

Demographic Variable	Component 1	Component 2
Economically Disadvantaged (ECD)	.252	.334
Ever English Learner (Ever EL)	.442	-.193
Underserved Race/Ethnicity (URE)	.427	-.057
Mobility	-.122	.866

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

Euclidean Distance

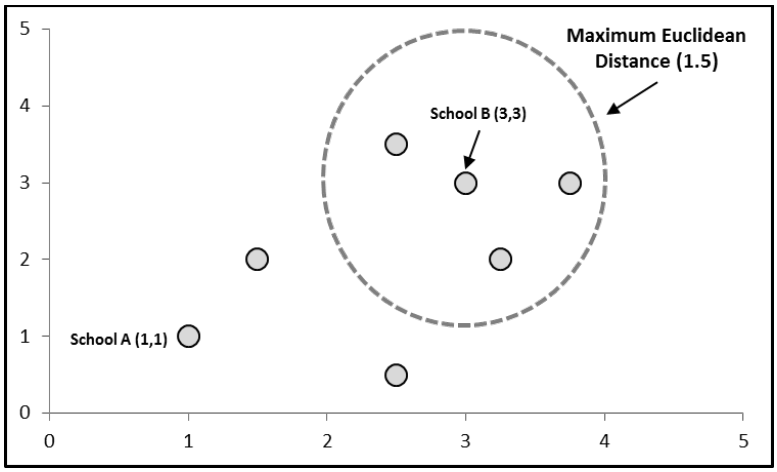
Euclidean distance is the distance between two points. The ODE weighted each school's component scores[6] and created a distance matrix to calculate the Euclidean distance for all pairs of schools within each school type. For example, the calculation of Euclidean distance between School A and School 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.



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

[6] The ODE weighted each component by its Eigen value. The Eigen value for the first component was 2.380 and the Eigen value for the second component was 1.065.

After calculating the Euclidean distance between all pairs of schools, the ODE removed all comparison schools where the Euclidean distance the focal school and comparison school was greater than 1.5. This considerably reduced the number of comparison schools 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.



Size Filter

The ODE applied a size filter to remove any comparison schools with considerable student enrollment size differences (i.e., sizes either substantially larger or smaller than the focal school). The size filter removed comparison schools with student enrollment sizes less than half or more than twice as large as the student enrollment size of the focal school. The table below 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 less than half or more than twice the size of School B.

Focal School Name	Comparison School Name	Euclidean Distance	Focal School Size	Comparison School Size	Size as a % of Focal School	Size 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	225	55.0	122.2	Yes
School B	School F	0.65	500	567	13.4	11.8	No
School B	School G	0.72	500	1,125	125.0	55.6	Yes
School B	School H	0.84	500	892	78.4	43.9	No

Twenty Most Similar Schools

Removing comparison schools using Euclidean distance and the size filter greatly decreased the number of comparison schools; however, while this left many schools with less than 20 comparison schools, a large number of schools (especially elementary schools) continued to have a large number of comparison schools. Thus, for focal

schools with greater than twenty comparison schools, the ODE selected the twenty comparison schools with the greatest proximity to the focal school as measured by the Euclidean distance. These final schools (twenty or less) are the comparison schools with the most similarity to the focal school according to student demographics and enrollment size. Lastly, these comparison schools were the basis for the calculation of the like-school averages on the school report cards.