

Nevada Peer Matching Tool Technical Documentation for Users

(Date last revised: October 12th, 2018)

Background: The Nevada Department of Education, in partnership with American Institutes for Research (AIR), has developed an online tool for schools to compare their results from the Nevada School Climate / Social Emotional Learning (NV-SCSEL) survey with those from similar schools throughout Nevada (“peer schools”). A matching algorithm, detailed below, automatically identifies peer schools based on key characteristics such as demographic characteristics of the students served and locale type.

Matching algorithm: Matches are based on the Mahalanobis distance, a widely used metric of standardized distance between points in multivariate space. For example, it has been used in diverse contexts such as [ecological modelling](#), [person identification software](#), [outlier detection](#), [chemical pattern recognition](#), and [DNA sequence dissimilarity](#), among many other applications.

This metric solves several challenges when comparing schools across multiple dimensions. For instance, raw differences in school characteristics cannot simply be summed because the scale differs for each characteristic (e.g., enrollment size varies more widely across schools than the percent of English Language Learners). The Mahalanobis distance metric solves this challenge by standardizing the variance for each school characteristic. In addition, it accounts for correlations among school characteristics, as [explained in detail elsewhere](#). Redundant information from highly related variables (e.g., percent proficient in math and percent proficient in reading) is not given undue weight because the Mahalanobis distance metric is based on the full covariance matrix. In other words, it is the multivariate generalization of standardized z-scores.

Data sources: Data on school characteristics came from the [Data Interaction for Nevada Accountability Portal](#) for the 2017-18 accountability year. NV-SCSEL survey results came from the spring 2018 NV-SCSEL survey administration.

Matching variables: For all schools, users can select the following characteristics as matching variables. Asterisked characteristics are selected by default:

- **Grade level*** (e.g., match a high school to only other high schools)
- **Race/Ethnicity:**
- **Race/Ethnicity:**
 - Percent American Indian or Alaska Native
 - Percent Asian
 - Percent Black/African American
 - Percent Hispanic*
 - Percent White
- **Special Populations:**
 - Percent Individualized Education Program*
 - Percent English Language Learners*
 - Percent Free/Reduced Priced Lunch Eligible*
- **Other characteristics:**
 - Enrollment size*

- Locale type (based on [NCES definitions](#)). If selected, only exact matches on major locale type are used (e.g., a rural school will match only to other rural schools). Those matches are additionally weighted by minor locale type (e.g., a rural remote school will match more often to other rural remote schools than rural fringe schools, holding all other characteristics constant). However, suburban schools can match with urban schools given the small number of suburban schools with spring 2018 NV-SCSEL survey data.

Grade-specific matching variables: If users select matches within grade level, they can also select matching variables that are specific to their school's grade level, which is the default option.

- **Elementary school:** Percent of students meeting the proficiency target on the CRT (New NV Standards) mathematics and reading assessments. Averaged across grades 3-5 (raw arithmetic mean).
- **Middle school:** Same as above but averaged across grades 7-8. Grade 6 data were not used because some elementary schools include grade 6.
- **High school:** Average ACT composite scores and 4-year graduation rates.

Missing data: Two strategies were used to handle missing school characteristics data:

- **Imputation:** In the source data on race/ethnicity and special populations, data are suppressed when there are fewer than 10 students in a category. For example, if there are more than zero but fewer than 10 Asian students in a school, the value of Percent Asian is suppressed (due to FERPA regulations). For schools with enrollment of 100 or more students, suppressed values were imputed as zero percent. This imputation rule reduces the frequency of errors in the matching algorithm, which requires complete data on all matching variables.
- **Restricted matches:** To handle remaining cases of missing data, matches are restricted to schools that have complete (i.e., non-missing) data for the variables used in the matching algorithm. In addition, those matching variables are restricted to those that have data for the user's chosen school (e.g., if Percent Asian is missing for the user's school, that variable is never used in the matching algorithm, even if the user clicks its checkbox).

Error checking: Using automated scripts, all possible combinations of selected schools and matching variables (over 200,000 combinations) were tested to systematically study the conditions under which the matching algorithm produces an adequate set of matches. This investigation found that the primary cause of errors was having too few potential eligible matches, usually due to missing data and simultaneously selecting both categorical variables (i.e., grade level and locale type). Having too few eligible matches is also a concern for match quality. To help ensure the quality of matches provided by the tool, users will receive an error message if the user options cause too few eligible matches which is defined as (a) fewer than 10 peer schools or (b) less than twice the number of requested matches. For instance, if there are 10 requested matches, there must be at least 20 eligible matches. This rule ensures that the matching algorithm is at least selecting the better half of eligible matches. The default number of matches is 5 schools.

For questions about the Nevada Peer Matching Tool, contact the NV-SCSEL survey team at nvschoolclimate@air.org