IES Institute of Education Sciences

A Guide to Identifying Similar Schools to Support School Improvement

Regional Educational Laboratory Central

At Marzano Research

REL 2021–096
U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Education Evaluation and Regional Assistance at IES

A Guide to Identifying Similar Schools to Support School Improvement

Douglas Van Dine, Bruce Randel, and Mary Klute July 2021

To support school improvement efforts, school leaders and education agencies might need to identify groups of schools that are similar so that schools can compare their performance or share practices with other schools in the same group. This could also allow education agencies to provide tailored supports to schools in a group. This guide describes how an education agency can select a distance measure (a statistical rather than a geographic measure) to identify schools that are similar to a target school, using a variety of characteristics that enable school leaders to better understand their schools' relative performance. This guide is based on work done with the Nebraska Department of Education and is designed to help staff in other education agencies who are interested in implementing a similar approach to support school improvement.

Contents

Why this guide?

As schools plan and evaluate improvement efforts, a key step is to identify groups of similar schools in order to make accurate and productive comparisons. School report cards produced by state education agencies often provide information not only on student test scores but also on student demographic characteristics, to facilitate the interpretation of academic outcomes in light of each school's student population. Research has demonstrated associations between the demographic characteristics of a student population (for example, the percentage of students living in poverty) and school performance (Hegedus, 2018), and education agencies have been urged to consider the student population of a school when assessing performance on academic outcomes (Darling-Hammond & Ascher, 1991; Hegedus, 2018; Salganik, 1994). Local education agencies can use this and other information to compare a school's performance to that of other schools with similar student populations. However, identifying similar schools and comparing them can be a complex undertaking (Salganik, 1994).

This guide presents an approach for identifying similar schools that is based on the one used by the Nebraska Department of Education (NDE). Research staff in other education agencies might consider the approach when charged with identifying groups of similar schools in their own districts to support school improvement (see appendix A for more information about the approaches used by other education agencies). Due to the technical aspects covered in this guide, the primary audience is research staff. The guide might also be useful for education leaders who need to decide how best to identify similar schools to meet their needs and fit their contexts.

This guide describes important decision points developed through collaboration between NDE and the Regional Educational Laboratory (REL) Central. NDE had found that Nebraska schools frequently compared their improvement efforts and outcome data (such as student achievement) against state averages or against nearby schools. Such comparisons are often misleading, however, because the schools could differ considerably in student demographic characteristics, enrollment, or other relevant factors. NDE recognized the importance of developing an approach for identifying groups of schools that are similar on a variety of characteristics to help local education leaders understand how their schools are performing. Additionally, NDE was interested in providing tailored support to schools with similar characteristics.

In 2018 and 2019 NDE explored approaches for identifying groups of similar schools. The REL Central supported NDE in exploring how to make decisions for identifying similar schools: which variables to include, what analytic method to use, and how to evaluate the quality of resulting matches. NDE wanted to incorporate a large and broad set of variables for schools to use in identifying similar schools for comparisons. And it wanted to encourage local education leaders to consider that schools similar to their own might not necessarily be located in neighboring districts.

This guide describes an approach based on the one NDE uses to identify similar schools. NDE developed the approach to meet the department's priorities: to consider many variables when calculating similarity and to identify the same number of similar schools for every school in the state. NDE wanted all schools to be able to connect with the same number of similar schools, to avoid sending the message that some schools were receiving more or fewer resources or that some schools had unique characteristics. Other education agencies might have different priorities and thus might prefer other approaches.

This guide can serve as a resource for research staff in other education agencies that are seeking guidance on how best to identify similar schools in their state to support their own school improvement initiatives. While developing the approach described in this guide, NDE discussed its efforts with other state education agencies and found several that were interested in learning more about the methodology in NDE's approach.

Overview of the Nebraska-based approach for identifying similar schools

Local education agencies can take any of several approaches for identifying groups of similar schools to support school improvement efforts. Some approaches apply methods that produce discrete groups of schools, with each school belonging to only one group. However, these methods can produce groups of similar schools that vary considerably in size, from just a few schools to many schools. For example, the New Mexico Public Education Department (2017) used latent class analysis to organize 847 schools into 14 groups. The largest group included 134 schools, and the smallest group included 9 schools (see appendix A for more information about the methods used by other education agencies). Such large variation can limit the meaningfulness of the information that a school might obtain from the group, especially if the school is unusual in its characteristics, including the students it serves.

Other approaches use a distance measure, which does not refer to geographic distance (such as miles) but to a statistical measure of how similar schools are (see box 1 for definitions of key terms). These methods use a set of variables to calculate a value for the difference—or distance—between each school and every other school in a state. An education agency can then use these values to determine how similar each pair of schools is on those variables. The smaller the value of the distance measure between two schools, the more similar the pair of schools. Using a distance measure allows for the identification of the same number of similar schools for each target school. Schools can be rank ordered by distance from the target school, and then a predetermined number or group size can be used to identify how many of those schools to include in the group of matched schools. For example, the Texas Education Agency uses a distance measure in its accountability system to identify a group of 40 similar schools, called comparison schools, for each school across the state (Division of Performance Reporting, 2017).

When examining approaches for identifying similar schools, the Nebraska Department of Education (NDE) initially considered cluster analysis and latent class analysis. However, as expected, these analyses resulted in large variations in the size of the resulting groups. NDE decided to look for a method that could control the number of matched schools so that leaders in each school would have the same, manageable number of matched schools for comparison.

NDE ultimately chose an approach that uses a distance measure because that method could accommodate more variables, allowed NDE to identify the same number of matched schools for every school in the state, and provided results that were easy for local education agencies to use and understand. In statistics, distance is often used as a summary measure of how similar two observations are, typically based on many variables at once. In this guide the distance between two schools is a measure of how similar they are.

Box 1. Key terms

Approach for identifying similar schools. All decision points in the overall process for identifying schools with similar characteristics.

Caliper. A threshold, or cutoff, identifying the maximum tolerated difference between matched schools.

Correlated variables. A measure of how variables “move” together; that is, how they relate to one another.

Distance measure. A measure of how similar two sample elements are, typically based on many variables at once. In this guide distance is used as a measure of how similar two schools are. A distance measure does not necessarily incorporate geographic distance. In this guide the two methods for calculating a distance measure are Euclidean and Mahalanobis (see appendix B).

Exact matching. A process in which matched schools have exactly the same value on the selected variable.

Index of similarity. A type of distance measure based on the measure or measures of interest, used to evaluate the quality of matches produced by a matching option.

Matched schools. A group of schools matched on a set of characteristics to an individual target school using a distance measure to assess similarity.

Matching option. The combination of the variables and the distance method used to identify groups of similar schools.

Matching with replacement. A process in which a school can be included in more than one group of similar schools. Once a school is identified as a match for one target school, it is placed back into the pool of schools and can be identified as a match for another target school.

Measure of interest. A variable on which schools need to be similar. This is used to evaluate the similarity of schools produced by a matching option.

Method. An established statistical or mathematical procedure.

Outlier. A value that is much smaller or larger than, or “lies outside,” most of the other values in a dataset. In this guide an outlier is a school that is dissimilar to other schools.

Similar schools. A group of schools that includes a target school and all identified matched schools.

Standardized variable. A variable that has been rescaled, typically to have a mean of 0 and a standard deviation of 1.

Stratified matching. A process in which matching is conducted separately within layers (strata) of schools, such as matching elementary schools to other elementary schools and matching middle schools to other middle schools.

Target school. An individual school to which other schools are matched.

Variable. A measure or characteristic on which schools may differ, such as number of enrolled students or percentage of students proficient in math.

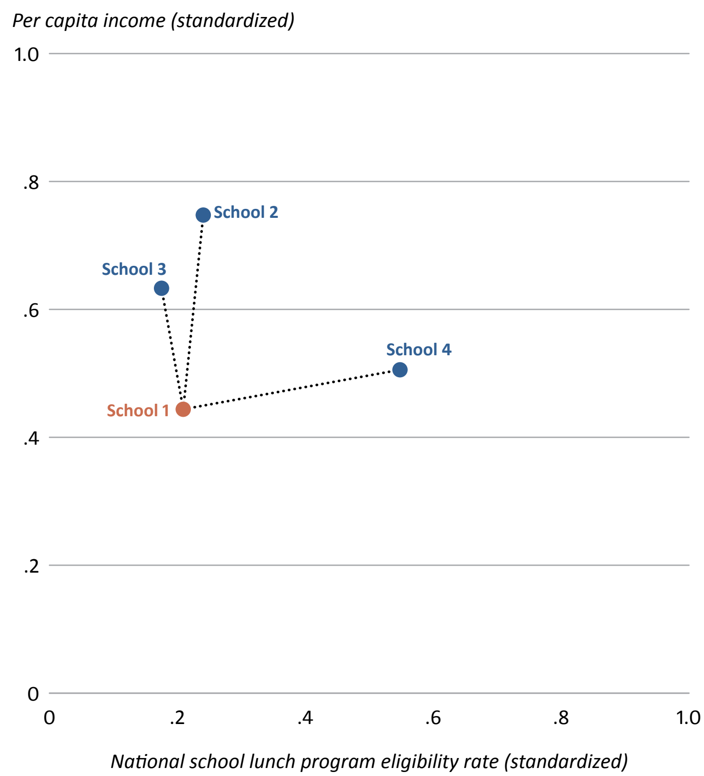
Variable scale. The unit or metric used to measure a variable, such as number of teachers, value in dollars, or percentage of students.

As other education agencies think about the Nebraska-based approach for identifying similar schools, it is important for them to reflect on the concept of distance and on what they hope to achieve by identifying similar schools. For illustration, consider an approach that uses two matching variables. First, schools are identified based on a set of variables selected to describe them. The set of variables locates each school at a specific point in a coordinate system. Suppose that there are four schools and two variables: percentage of students eligible for the national school lunch program (national school lunch program eligibility rate) and per capita income for the school district. The location of each school can be plotted on a two-dimensional graph (figure 1). Because of differences in scale for these two variables (percentages and tens of thousands of dollars), values are standardized.

Next, after all schools have been located as points on the graph, the distance between pairs of schools can be calculated. The distance can be described as the shortest straight-line path connecting the two points identifying a target school and one of its potential matched schools. For example, in figure 1 school 1 is the target school, and schools 2, 3, and 4 are potential matched schools. As the dashed lines show, school 1 is closer (or more similar) on these two variables to school 3 than it is to school 2 or school 4.

If there are three matching variables, the locations of schools can be visualized as points in a three-dimensional coordinate system. Locations are not easily visualized or conceptualized beyond three dimensions, but the values for a given set of variables can still be

Figure 1. Example of plotting the distance between four schools based on two variables



Source: Authors' construction.

used to identify the location of each school as a specific point within that multidimensional coordinate system. The locations of these points (schools) can then be used to calculate the distance between each pair of schools.

These calculations yield a unique set of matched schools for each target school (in this case school 1). NDE decided to use matching with replacement, meaning that a school that is identified as a match for a target school is placed back into the pool of schools and can be identified as a match for another target school. As a consequence, a school can serve as a match for many other schools, and the resulting groups of identified similar schools are overlapping, with a given school potentially appearing in multiple groups of identified similar schools. Alternatively, an education agency could decide to use matching without replacement, meaning that a school that is identified as a match for a target school cannot be matched to another target school, so the resulting groups would not overlap. Using matching without replacement, however, could reduce the quality of the matches. For example, suppose that a matched school is identified as the most similar school to two target schools. If matching without replacement is used, the matched school can be paired with only one of the target schools, so the group of similar schools for the other target school will not include that school's closest match.

The Nebraska-based approach for identifying similar schools

This guide outlines an approach that resulted from collaboration between the Nebraska Department of Education (NDE) and the REL Central, and thus the approach is tailored to NDE's needs. While describing the approach, the guide also highlights the decisions that NDE made in developing the approach and the reasons for those decisions (figure 2). The Nebraska-based approach is offered as an example for education agencies to consider as they seek to identify the approach that best fits their own contexts and needs.

Figure 2. Steps and decision points in the Nebraska-based approach for identifying similar schools

Identify a set of variables that meets key considerations			
What is the intended goal of identifying similar schools?	Which variable categories and variables should be included?	Is an exact match preferred for some variables?	Should averages of variables across multiple years be included?

↓

Choose a suitable distance measure and calculate the distance between every pair of schools		
What distance measure should be used?	What happens when variables have different scales?	What happens when variables are correlated?

↓

Match schools to each target school		
Is a predetermined number of identified matches desired?	Should identified matched schools be limited to those within a maximum distance from the target school?	What is the difference between choosing a predetermined number and setting a maximum distance?

↓

Evaluate the quality of matches produced by a matching option
Which combinations of variables and distance measure result in the best matches?

Source: Authors' construction.

Identify a set of variables that meets key considerations

The Nebraska-based approach for identifying similar schools can accommodate a large number of variables, so education agencies using this approach need to decide which variables to include. Agencies can consider options and decision points in four areas:

1. What is the intended goal of identifying similar schools?
2. Which variable categories and variables should be included?
3. Is an exact match preferred for some variables?
4. Should averages of data across multiple years be used?

What is the intended goal of identifying similar schools?

Education agencies should consider the intended goal of identifying similar schools when selecting variable categories and variables. For example, as agencies decide whether to include variables in the category of student achievement, they might consider the following options:

- If the goal is to evaluate how schools are performing relative to similar schools, it would not be appropriate to include current outcome data, such as student proficiency rates or graduation rates, when calculating the distance measure. Doing so would result in schools being matched to other schools with similar student achievement levels.[1]

- If the goal is to create opportunities for schools to connect and share practices with one another, the decision about whether to include current outcome data is less clear.

    - If current outcome data are not included when calculating the distance measure, the identification of similar schools can create an opportunity for a school struggling in one area (for example, graduation rates) to connect with and learn from other schools that serve similar populations and are excelling in that area.

    - If current outcome data are included, the identified similar schools will likely face similar challenges and can collaborate to explore possible solutions.

- If the goal is to help an education agency tailor its supports to schools, the agency might consider including a school's accountability level or student outcomes when calculating the distance measure. For example, the agency might identify a group of schools that serve similar populations of students, have low student achievement, and struggle with student attendance. This group of schools might benefit from state supports that differ from those provided to a group of schools with low student achievement and high attendance rates.

1. In some instances it might be useful to include outcome data from previous years (for example, outcomes from previous years for current students or outcomes from previous cohorts for students in the same school) when identifying similar schools. Controlling for outcomes from previous years for current students produces an estimate of a school's current performance. Controlling for outcomes from previous cohorts produces an estimate of change in a school's performance over time.

NDE's goal in identifying similar schools was to develop tailored support for similar schools within accountability levels. This goal drove the decision to include student outcomes (for example, current percentages of students proficient in English language arts and math) when calculating its distance measure.

Which variable categories and variables should be included?

When selecting variables to include, education agencies might start by considering which categories of variables to include. Categories typically reflect key clusters of variables central to the intended goal; for example, test scores in particular subject areas, behavioral outcomes, or measures of classroom practices (Schochet, 2008). Education agencies might want to consider variables that influence school quality or student success (for example, attendance rate, suspension rate), as well as student outcome variables (for example, proficiency in English language arts or math).

Education agencies might consider categorizing variables using a format similar to that of a logic model, including "inputs" that are beyond the immediate control of a school, "policies and practices," and "outcomes." For example, student background and community demographic characteristics are beyond a school's control. Staff characteristics and school-level expenditures could be considered school practices because administrators have some control over them. Student behavior or discipline and student achievement could be outcomes of interest. These are only examples, however. In some contexts variables such as school-level expenditures might not be within a school's control. Additionally, student academic outcomes can be influenced by student behavior or discipline, so an additional category of "intermediate outcomes" might be included.

Once the categories of variables are identified, education agencies might consider the following factors when deciding which variables to include in the identified categories:

- If only variables beyond the control of a school are used to identify matches, differences in outcomes within matched schools can be interpreted as capturing differences in performance.
    - In some cases variables that capture school practices might be beyond a school's control, but in other cases they might not be (for example, school-level expenditures). Education agencies should consider the context and goal of identifying similar schools when determining whether to include these variables to identify matches.
- If outcomes are used as control variables when identifying matches, differences in outcomes within matched schools cannot be interpreted as capturing differences in performance.
    - If education agencies include outcomes to identify matches, interpreting the remaining differences becomes difficult. Although matched schools might be able to share practices with one another, comparisons within each set of matched schools could be misleading. For example, if education agencies match on outcomes, it will not be clear which schools outperform others because schools will also be grouped by similar performance levels.

Box 2 offers examples of how some education agencies have selected which variables to include.

Box 2. What rationales have other education agencies used for selecting variable categories?

Student characteristics. In previous applications of approaches for identifying similar schools, some education agencies have focused on the variables category of student characteristics. For example, when identifying inputs for the School Efficiency Metric, the U.K. Department for Education (2018) includes only measures within a school's control and considers the extent to which schools contribute to student achievement. The department views per pupil funding as representative of all inputs within a school's control. When identifying comparison schools, the Oregon Department of Education (n.d.) relies on four student characteristics: national school lunch program eligibility rate, percentage of students ever identified as English learner students, percentage of students belonging to an underserved racial/ethnic group, and percentage of students mobile within the school year.

School and student characteristics. Other education agencies include the variables category of school characteristics. For example, the Texas Accountability System identifies each school campus first by school type—elementary, elementary/secondary (K–12), middle school, or high school (Division of Performance Reporting, 2017). The campus is then grouped with 40 other campuses, which can be located anywhere in Texas, that are identified as most similar using measures of both school characteristics (for example, grade levels served, student population) and student characteristics (for example, national school lunch program eligibility rate, mobility rate, and percentage of English language learner students).

See appendix A for more information about the approaches used by these and other education agencies.

Understanding that a broad range of factors is related to school quality and success, NDE considered variable categories that included student, staff, school, district, and community characteristics (Office of Data, Research and Evaluation, 2019). NDE identified a set of 27 variables in six categories that were relevant, available for all schools in the state, and persistent (that is, regularly and consistently collected) to describe any given school (table 1).

Based on recent research suggesting that matching within a local area might improve the matching for unobserved characteristics (see, for example, Cook et al., 2020), education agencies might want to consider adding a variable for geographic location. NDE chose to consider geographic location separately rather than including it in its set of variables to identify similar schools because education leaders in Nebraska often already looked to nearby schools or schools within the same athletic conference as comparison schools. These education leaders tended to consider a school's geographic proximity or comparable student population to be more important than other characteristics that might identify better matched schools. NDE wanted to encourage school leaders to consider that more similar schools might exist outside their local areas and that these schools could be identified by characteristics beyond student population. Even so, NDE developed its website so that users can compare a target school either with schools that are geographically closest to that school or to a group of 12 similar schools based on the 27 identified variables (see table 1).

Table 1. Variables used in the Nebraska-based approach for identifying similar schools, by variables category

Variable category	Variable
Student characteristics	EL rate Homeless rate Migrant rate Minority rate NSLP-eligibility rate
School characteristics	Enrollment
Student performance	Attendance rate ELA percent proficient Graduation rate Math percent proficient Science percent proficient
Staff characteristics	Average years of teaching experience Percentage of teachers with master's degree
Student behavior/discipline	Unduplicated expulsions Unduplicated suspensions
School funding	Grand total of all receipts Per pupil cost by average daily membership[a]
Community demographic characteristics	Gini Index[b] Land area Labor force participation rate Land valuation Median household income Per capita income Percentage 25+ with bachelor's degree or more Population density Total population Unemployment rate

Note: The variable names in this table are those used by the Nebraska Department of Education for identifying similar schools and in some cases do not precisely match the terminology used elsewhere in this report.

a. Total annual costs divided by the average daily membership for the district. The average daily membership calculation aggregates the numbers of days in session for all students. The number of days in session for each student is determined by looking at the calendar track to which the student is assigned and summing the number of in-session days reported under that track. Only in-session days that are between the student's enrollment entry and exit dates are counted.

b. A summary measure of income inequality, ranging from 0 (perfect equality, or everyone receives an equal share) to 1 (perfect inequality, or only one person or group receives all the income; United States Census Bureau, 2016).

Source: Office of Data, Research and Evaluation, 2019.

Is an exact match preferred for some variables?

When selecting variables, education agencies might identify specific variables that schools must match on exactly to be considered similar. In such cases schools are first organized into groups based on these variables, which are typically called strata. Next, the distance measure is calculated for all pairs of schools within each stratum. This is referred to as stratified matching.

For example, education agencies might want to match exactly on the grade span of schools. Doing so would ensure that all elementary schools are matched to other elementary schools, and so forth. Exact matching on grade span might be more challenging in states that commonly have schools that cross the traditional grade spans of elementary, middle, and high, such as K–8 schools and middle-high schools. In this case education agencies might consider identifying separate sets of matched schools for each grade span at a school. For example, a K–8 school might have one set of matched schools for grades K–5 and another set of matched schools for grades 6–8.

When considering variables for exact matching, education agencies should keep in mind their intended goals for identifying similar schools. For example, agencies might consider the following options:

- If the goal is to evaluate how schools are performing relative to similar schools, it might be useful to identify similar schools within grade spans because accountability measures often vary substantially by grade level. For example, high school graduation rates are often included in accountability ratings for high schools, but this indicator does not apply to elementary schools. Matching schools within grade spans will ensure that all similar schools are assessed using the same accountability measures as the target school.

- If the goal is to create opportunities for schools to connect and share practices with one another and if in-person connection is deemed important, it might be useful to identify similar schools within geographic regions.

- If the goal is to help an education agency tailor its supports to schools, it might be useful to identify similar schools within accountability levels. For example, Nebraska has four accountability levels: excellent, great, good, and needs improvement.

When selecting variables for exact matching (strata), education agencies should pay close attention to cell size—the number of schools in each of the layers defined by the set of strata. It could be tempting to select many variables for exact matching, such as school locale, grade span, student enrollment, and geographic region. However, attempting to exactly match on too many variables at once may result in very small cell sizes, which will limit the utility of the overall approach for identifying similar schools. For example, a specific region in a state might include only two rural schools that serve grades K–5 and have a total enrollment of 200 students. When selecting variables for exact matching, education agencies might first calculate the number of schools in each group and consider the desired number of matched schools to be identified for each school. For example, if the goal is for each target school to have at least four peer schools, exact matches that provide only two schools cannot be included.

Should averages of variables across multiple years be used?

Education agencies might also find merit in using multiple years of data (for example, data from prior years for the current cohort of students or data for previous cohorts in the same school) to reduce the amount of change from year to year. Because school data are often collected and examined on an annual basis, agencies might be tempted to conduct analyses annually to identify similar schools based on the most recent data available. One potential pitfall in this approach is that identified similar schools can change substantially from one year to the next, depending on the variables selected, thereby diminishing the usefulness of identifying similar schools.

Regardless of the goals of identifying similar schools, education agencies might prefer to have some stability from year to year in the set of similar schools. Agencies might consider doing this for the following reasons:

- If the goal is to evaluate how schools are performing relative to similar schools, school leaders might have difficulty understanding their school's progress if they are comparing their school to substantially different schools each year.
- If the goal is to create opportunities for schools to connect and share practices with one another, school leaders might struggle to build relationships with leaders from identified similar schools if the group of similar schools changes substantially from year to year.
- If the goal is to help an education agency tailor its supports to schools, the process of designing and implementing that support could take several years, and changing schools during that time might cause confusion.

One technique to guard against substantial changes from year to year is to base calculations on moving averages. For instance, when data are available on an annual basis (for example, student enrollment, attendance rate, percentage of teachers with at least a master's degree), education agencies might use an average of the last two or three years of data rather than only the current year of data. Agencies might also consider identifying similar schools every two or three years instead of annually.

Choose a suitable distance measure and calculate the distance between every pair of schools

After education agencies identify the variables they want to include, they should determine which method to use to calculate the distance measure used to determine how far apart two schools are from each other in a space defined by variables. The shorter the distance between two schools, the more alike they are. This section provides guidance for choosing between two common methods of calculating a distance measure: the Euclidean distance method and the Mahalanobis distance method. Each method has advantages and disadvantages (see appendix B for more details about the two methods).

What distance measure should be used?

Euclidean distance method

In general, using the Euclidean distance method is simple. It involves computing the ordinary straight-line distance between two schools (or points) using a formula such as $d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. These calculations involve two variables, x and y, and the distance can be drawn as the straight line between the two points, as in figure 1.

As an illustration, consider school 1 and school 2 from figure 1. School 1 is located at $x_1 = 0.21$ and $y_1 = 0.44$, and school 2 is located at $x_2 = 0.24$ and $y_2 = 0.75$. The Euclidean distance between school 1 and school 2 would be $\sqrt{(0.21 - 0.24)^2 + (0.44 - 0.75)^2}$, or 0.31. Calculating and visualizing Euclidean distance become more complicated when many variables are included in the calculation because of a limited ability to conceptually understand space and distance beyond three dimensions (see appendix B for how to calculate Euclidean distance with more than two variables). Including more variables often results in introducing variables with different units, such as per capita income and attendance rate, or variables that are correlated with one another, such per capita income and percentage of students eligible for the national school lunch program.

By design, the Euclidean distance method allows all variables to contribute equally to the distance measure, which gives rise to two primary issues when identifying matches. First, Euclidean distances are heavily influenced by the scales of the variables selected. When variables have different scales, the computed distance measure tends to be dominated by or skewed toward the variables with the largest scales. For example, the difference in scales between per capita income (measured in tens of thousands of dollars) and high school graduation rate (measured as a percentage) means that per capita income would dominate high school graduation rate in identifying how similar schools are. To neutralize the effects of different variable scales, it is common to standardize variables before using the Euclidean distance method.

Second, the Euclidean distance method can give too much weight to correlated variables if they are capturing a single underlying trait. Because each variable contributes equally in the calculation of the Euclidean distance measure, the use of highly correlated variables can skew the distances toward those variables even if they are measuring a single trait. In other words, correlated variables potentially replicate information (or introduce redundant information) so that calculated Euclidean distance measures are skewed along the line of correlation.

For example, suppose that the variables included for calculating distance are gender, English language proficiency, and English language arts scores, and suppose that gender is weakly correlated with English language proficiency and English language arts scores. In this case school leaders would probably decide to retain the gender variable but would then need to decide whether to include both English language proficiency and English language arts scores. If those scores are not correlated, both variables would likely be included because they are capturing different characteristics. However, if English language proficiency and English language arts scores are highly correlated, school leaders might choose to drop one of them because they are likely measuring some similar, underlying, unobserved trait. If all three variables are retained and if English language proficiency and English language arts scores are capturing a single trait, then the Euclidean method would give too much weight to that trait.

Mahalanobis distance method

The Mahalanobis distance method addresses both issues with the Euclidean distance method described above: it standardizes the variables so that each has a variance equal to one, and it transforms the variables to remove any correlation. Thus, the Mahalanobis distance method implicitly downweights any difference in scale or correlated variables (see appendix B for how to calculate Mahalanobis distance).

What happens when variables have different scales?

When including variables with different scales in calculations to identify similar schools, education agencies often take the additional step of standardizing the variables before using the Euclidean distance method. The Mahalanobis distance method automatically standardizes the variables.

To understand why differences in the scale of variables are important, consider the three schools and four variables shown in table 2. Per capita income is measured on a much larger scale (tens of thousands of dollars) than the other three variables; the national school lunch program eligibility rate and attendance rate are percentages, and enrollment is generally in the hundreds or thousands. In the example, community per capita income is similar for schools 1 and 7 but much lower for school 3. On the three other variables schools 1 and 3 are similar, but school 7 is different.

The Euclidean distance method for calculating unstandardized distance measures between each pair of schools produces the following results:

- The Euclidean distance between schools 1 and 3 is 26,846.85.
- The Euclidean distance between schools 1 and 7 is 1,201.82.
- The Euclidean distance between schools 3 and 7 is 28,008.18.

Thus, without standardization of the variables, schools 1 and 7 would be identified as the most similar because the distance between them is the smallest. However, schools 1 and 7 are similar on only one of the four variables: per capita income. This variable dominates the calculation because of its much larger scale, which makes the schools appear more similar than they likely are.

The example in table 2 shows that schools 1 and 3 are similar on three of the four variables, whereas schools 3 and 7 are different on three of the four variables. However, when the unstandardized Euclidean distance measure is calculated, the distance between schools 1 and 3 is quite similar to the distance between schools 3 and 7. Even though these two pairs are different, the distances appear similar because the difference in per capita income is similar for each pair.

Table 2. Example for three schools and four variables

School	National school lunch program eligibility rate (percent)	Per capita income (dollars)	Enrollment (number of students)	Attendance rate (percent)
1	32.4	44,116.28	364	94.7
3	34.4	17,269.44	363	94.2
7	2.5	45,275.85	50	99.7

Source: Authors' construction.

If these three variables are first standardized so that each has a mean of 0 and a standard deviation of 1:

- The Euclidean distance between schools 1 and 3 is 1.71.
- The Euclidean distance between schools 1 and 7 is 2.92.
- The Euclidean distance between schools 3 and 7 is 3.55.

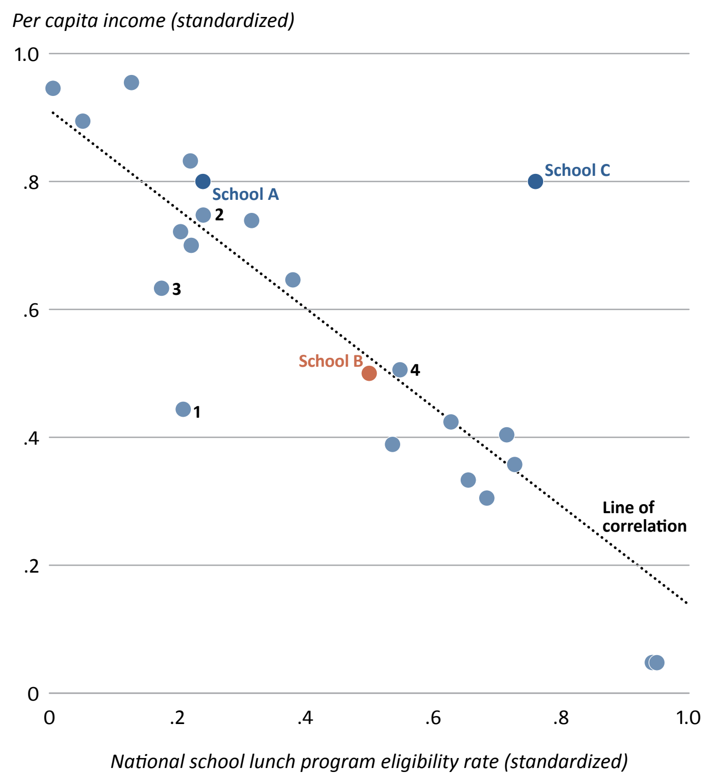
Thus, schools 1 and 3 would be identified as the most similar.

What happens when variables are correlated?

Correlated variables are another issue with the Euclidean distance method, whereas the Mahalanobis distance method inherently accounts for any correlation.

To better understand why correlation is an issue, consider the plot of data points representing the national school lunch program eligibility rate and per capita income (figure 3). This plot builds on the plot in figure 2, adding more nearby schools (the points labeled 1, 2, 3, and 4 represent the schools shown in figure 2). The plot reveals a positive correlation (indicated by the dotted line) between the two variables: a higher national school lunch program eligibility rate is negatively correlated with per capita income.

Figure 3. Example of plotting data points for multiple schools showing correlation between two variables

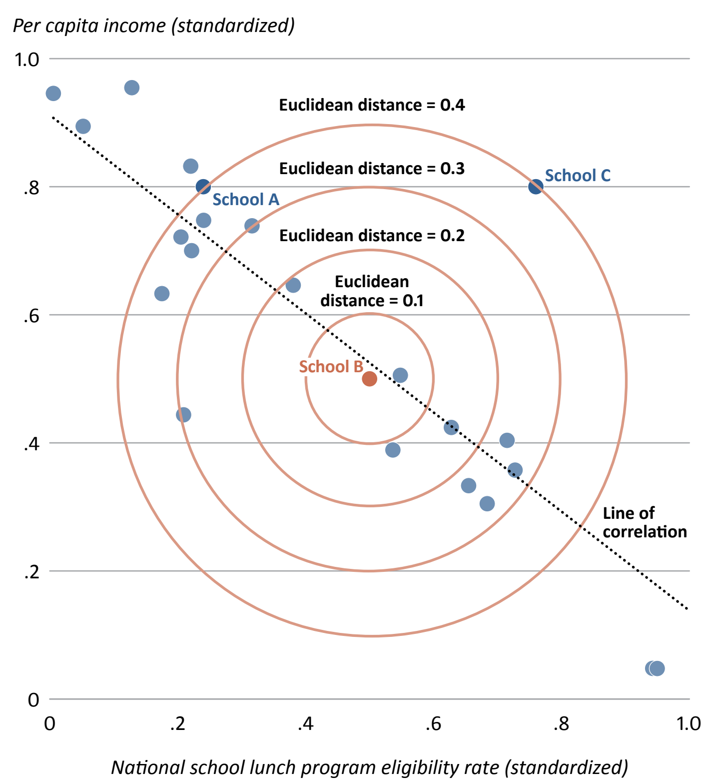


Source: Authors' construction.

The Euclidean distance method ignores correlation and calculates straight-line distance. Euclidean distance is often represented as concentric circles around a point of interest (figure 4). For example, with school B as the point of interest or target school, the concentric circles represent Euclidean distance measures of 0.1, 0.2, 0.3, and 0.4 from target school B. Thus, in the Euclidean distance method, schools A and C are equally distant from school B, as they are both 0.4 from school B. That is, the Euclidean distance between schools A and B is the same as that between schools B and C. The Euclidean distance method could still be appropriate if there are strong prior indications that the variables represent distinct traits that should be given equal weight relative to other variables in the model, regardless of any correlations.

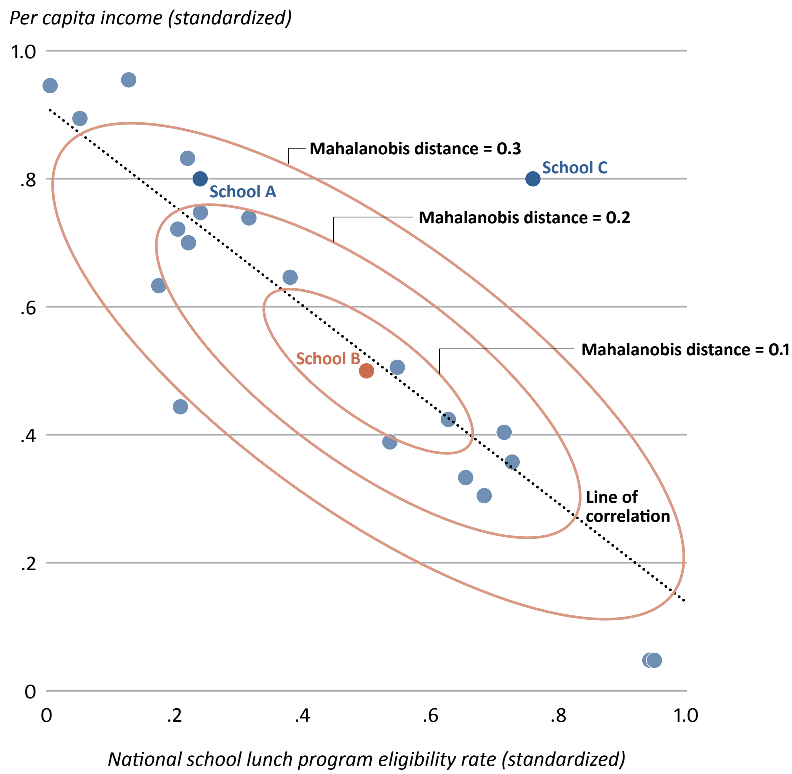
When the Mahalanobis distance method is used, any correlation is accounted for in the calculation. Mahalanobis distance is often represented as ovals stretched along the line of correlation and centered on a point (in this case, a school) of interest (figure 5). The ovals represent Mahalanobis distance measures of 0.1, 0.2, and 0.3 from school B, the target school. Recall that with the Euclidean distance method schools A and C are both 0.4 from school B (see figure 4). With the Mahalanobis distance method, however, school A is less than 0.3 from school B, whereas school C is more than 0.3 from school B. Another way to think of this is that school C is an outlier in the space defined by the national school lunch program eligibility rate and per capita income because of its distance from the line of

Figure 4. Example of Euclidean distance measure with correlated variables and with school B as the target school



Source: Authors' construction.

Figure 5. Example of Mahalanobis distance measure with correlated variables and with school B as the target school



Source: Authors' construction.

correlation. Schools A and B both show the expected relationship between national school lunch program eligibility rate and per capita income, meaning that they are more similar to each other than they are to school C. In cases such as these, when variables are correlated and might be capturing a single underlying trait, the Mahalanobis distance method might be more appropriate.

The Nebraska Department of Education's selection of a distance measure

NDE decided to use the Euclidean distance method in identifying similar schools in Nebraska. This decision was due partly to the simplicity of calculating Euclidean distances. Like other education agencies that have used the Euclidean distance method (Division of Performance Reporting, 2017; Oregon Department of Education, n.d.; U.K. Department for Education, 2018), NDE accounted for differences in variable scales by standardizing the variables before calculating the distance measures (see appendix A for examples of how some other education agencies have incorporated standardization). NDE also determined that the correlation between its included variables was low, so it had few concerns that the results would be skewed.

Match schools to each target school

Once distance measures are calculated between every pair of schools, an education agency can use these measures to identify a group of matched schools for each target school. First, the matched schools are rank ordered by distance from the target school, from most similar to least similar. Then, if the education agency wants the group of matched schools to be smaller than the entire sample of matched schools in the ordered list, the agency can determine the group of matched schools in one of two ways: choosing a predetermined number or setting a maximum distance.

Is a predetermined number of identified matches desired?

One common way to reduce the size of the group of matched schools is to select a predetermined number of schools to include in the group. The education agency determines how many schools it wants to include as matches to the target school and, from the ordered list, includes only schools up to that threshold rank so that every target school has the same number of matched schools. In this way schools that are considered outliers have the same number of matched schools as schools that are closer to the mean do. Additionally, when the goal of identifying similar schools is to create opportunities for schools to connect, a predetermined number of matched schools provides each school with an equal number of other schools with which to potentially connect. Other ways to identify sets of matched schools often make it difficult for outlier schools to find matched schools or for schools to meaningfully connect.

One potential disadvantage of this way to identify sets of matched schools is that the quality of the matches is likely to vary across schools. Some schools could have matched schools that are similar to them, but other schools, especially outlier schools, could have matched schools that are dissimilar to them.

Should identified matched schools be limited to those within a maximum distance from the target school?

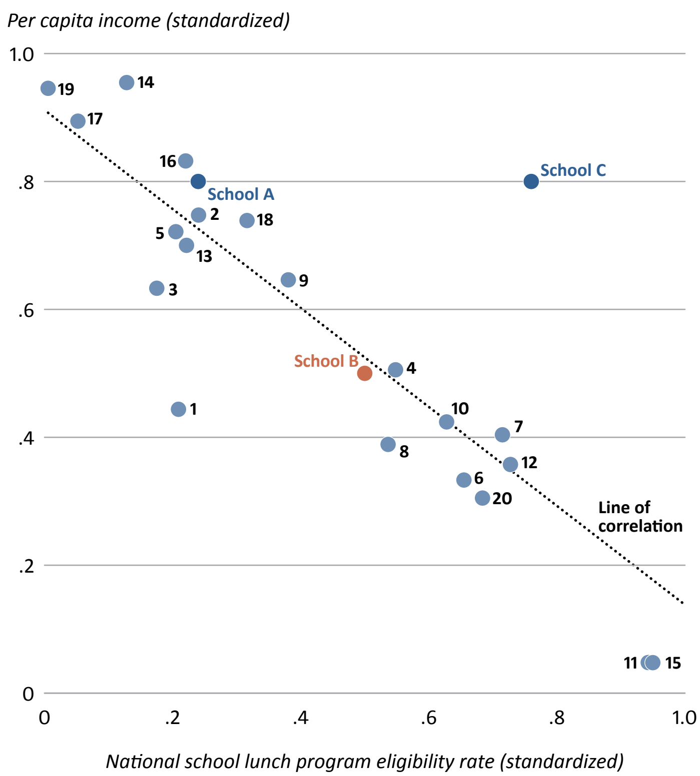
An alternative way to identify groups of similar schools is by selecting a value on the distance measure, called a caliper, which establishes the maximum distance that a target school and another school can have from each other and still be considered similar. All schools with distance measures greater than the caliper are not considered similar schools. A caution is that using a caliper could result in fewer schools being matched to outlier schools in the state. An advantage of using a caliper is that it helps ensure uniform similarity among groups of similar schools. Using a caliper eliminates the possibility that a target school has outlier matched schools that are actually dissimilar to the target school based on the variables used to calculate the distance measure.

What is the difference between choosing a predetermined number and setting a maximum distance?

Education agencies need to decide on the best way to identify sets of matched schools, depending on their contexts and needs. The following example can help clarify the difference between choosing a predetermined number of matches and setting a maximum distance (or caliper). Figure 6 plots the two variables for 23 schools used in previous examples: national school lunch program eligibility rate and per capita income. Because these two variables have different scales, they were standardized before plotting the locations of the schools. Because the two variables are negatively correlated, it might make sense to use the Mahalanobis distance method to accommodate for this negative correlation. However, this example uses the Euclidean distance method.

With schools A, B, and C identified as individual target schools, the distances are calculated between each target school and each of the other 22 schools. The 10 most similar (closest) schools to each target school can then be listed and organized by distance from the target school (table 3). For example, school 16 is 0.04 from school A and is therefore the most similar to school A. Note that school B, which is another target school, is included in the group of similar schools for school C. This highlights that, when using matching with replacement, every school in the state is considered a potential match for any target school.

Figure 6. Example of plotting 23 schools using two variables



Source: Authors' construction.

Table 3. The 10 most similar schools for each of three target schools, by Euclidean distance

School A		School B		School C	
Similar school	Euclidean distance	Similar school	Euclidean distance	Similar school	Euclidean distance
School 16	0.04	School 4	0.05	School 4	0.37
School 2	0.05	School 8	0.12	School 7	0.40
School 5	0.09	School 10	0.15	School 10	0.40
School 18	0.10	School 9	0.19	School B	0.40
School 13	0.10	School 6	0.23	School 9	0.41
School 3	0.18	School 7	0.24	School 12	0.44
School 14	0.19	School 20	0.27	School 18	0.45
School 9	0.21	School 12	0.27	School 8	0.47
School 17	0.21	School 1	0.30	School 6	0.48
School 19	0.28	School 18	0.30	School 20	0.50

Source: Authors' construction.

However, in addition to identifying similar schools, an education agency might want to minimize the dissimilarity among matched schools. The agency could decide that a matched school would need to be within a maximum distance from a target school. For example, the agency might set the maximum allowable distance at 0.20. In this case school A would have seven identified matched schools (schools 16, 2, 5, 18, 13, 3, and 14), school B would have four (schools 4, 8, 10, and 9), and school C would have none. Thus, each school would end up with a different number of similar schools. If using the selected maximum allowable distance resulted in the creation of groups of similar schools that were different in size, the agency might decide to adjust the maximum allowable distance so that groups are the same size. In the previous example, if the maximum allowable distance were increased to 0.30, schools A and B would each have 10 matched schools. However, school C would still not have any matched schools because it is an outlier.

Choosing a predetermined number of matched schools and setting a maximum distance can produce different results, so deciding which to use depends on the context and purposes of the education agency. In some cases, however, using both might make sense. For example, an education agency might initially set a maximum distance that results in large sets of matched schools for some target schools. In this case, to ensure that the sizes of groups of similar schools are useful for target schools, the agency might additionally decide to select a predetermined number of matched schools.

The Nebraska Department of Education's selection of matched schools

NDE's priority was to provide the same number of matched schools for each target school in the state. Due to the variation in the characteristics of schools across the state, NDE identified the 12 most similar schools (schools with the smallest distance measures) for each target school on all 27 identified variables.

Evaluate the quality of matches produced by a matching option

As education agencies consider the decision points described in the previous section, they might want to evaluate and compare results from different matching options. That can be done by using different sets of variables, here considered as different distance measures, to identify groups of similar schools. One process for evaluating the quality of matches produced by a matching option draws on methods used to evaluate balance (that is, baseline equivalence) between groups produced by matching, often the creation of a treatment group and a comparison group (Luellen et al., 2005; Stuart, 2010; box 3). Education agencies can use this evaluation process for the variables on which schools need to be most similar.

Box 3. Steps in evaluating the quality of matches

Step 1. Select the measures of interest.

Step 2. Calculate the mean and standard deviation of the matched schools.

Step 3. Calculate the difference between each school and its matched set.

Step 4. Summarize the differences.

Step 5. Summarize the variation among matched schools.

Step 6. Identify and count the outliers.

Step 7. Repeat steps 2–6 for all matching methods and measures of interest.

Step 8. Evaluate the matching options.

Step 1. Select the measures of interest

Because the purpose is to evaluate matching options by estimating both how close groups for each target school are and how close schools within individual groups are, it is not necessary, or even advisable, to include all the variables used to identify similar schools. Use only the most important variables on which the schools need to be similar to evaluate the quality of matches. If the matching variables can be grouped into categories, use one or two of the most important or most representative variables from each category as the measures of interest.

The evaluation can be conducted separately for each measure of interest, regardless of how many variables are used in calculating the distance measure. As recommended above, standardize the variables used in matching before calculating the distance measures. Variables from the previous school year can also be used in evaluating the matching.

Step 2. Calculate the mean and standard deviation of the matched schools

Calculate the mean and standard deviation of schools matched to each target school on the measures of interest. Calculate these values separately for each measure of interest. The standard deviation of each group of matched schools, which is a measure of variation, provides an index of how similar the schools are within the group.

For example, suppose that for a target school with 12 matched schools the standardized mean for the national school lunch program eligibility rate is 1.59 with a standard deviation of 0.79.

Step 3. Calculate the difference between each school and its matched set

Compare each target school to its matched schools by calculating the difference between its value on the measures of interest and the standardized mean of its matched schools (from step 2). This calculation is done by subtracting the mean of the matched schools from the target school's value to estimate how close the matched schools are to the target school.

Step 3 can be calculated as:

$$d_i = x_i - \overline{x_g}$$

where d_i is the difference between school *i* and its matched group mean, x_i is the value on the variable of interest for school *i*, and $\overline{x_g}$ is the mean on the measure of interest for school *i*'s matched group.

Continuing with the example from step 2, the standardized national school lunch program eligibility rate of the target school is 1.95. Thus, the target school value minus the matched school mean yields a difference of 0.36 (that is, 1.95 − 1.59 = 0.36).

Step 4. Summarize the differences

For each measure of interest, summarize the differences from step 3 across the entire population of schools by calculating the average of the absolute values of those differences. This average provides an overall index of similarity for how close the target schools are to their matched schools on the measures of interest.

Step 4 can be calculated as:

$$D_p = (|d_1| + |d_2| + |d_3| + \ldots + |d_n|)/n$$

where D_p is the population average of the absolute differences, $|d_1|$ to $|d_n|$ are the absolute values of the differences calculated in step 3 for school 1 to school n across the population of schools, and n is the number of matched groups in the population.

In the example, across the entire population of schools, the mean of the absolute value of the differences in national school lunch program eligibility rate is 0.02.

Step 5. Summarize the variation among matched schools

For each measure of interest summarize the variation (from step 2) of matched schools across the entire population by calculating the mean of the standard deviations of all the groups of matched schools. This average provides an index across the entire population for how close the matched schools are to other schools in their groups.

Step 5 can be calculated as:

$$M_{SD} = (SD_1 + SD_2 + SD_3 + \ldots + SD_n)/n$$

where M_{SD} is the mean of the standard deviations across all the matched groups in the population, SD_1 to SD_n are the standard deviations from step 2 of all the groups, and n is the number of matched groups in the population.

In the example the mean standard deviation in national school lunch program eligibility rate is 0.17 across the entire population of schools.

An alternative is to calculate the square root of the average variance for each group, where group variances are their standard deviations squared.

Step 6. Identify and count the outliers

Outliers are schools in a matched group that have a value on a measure of interest that is greater than two standard deviations from that group's mean,[2] using the standard deviation

2. The literature describes many ways to define outliers, and each definition suits a different purpose. A common purpose is to identify and remove outliers from a dataset; for this, the definition of an outlier is typically 1.75 × interquartile range. The purpose here is to provide a simple way (standard deviation) to identify outliers using descriptive statistics that are easily calculated, as one way to evaluate the results produced by different matching methods. This simpler method should suffice for this purpose.

of the matched group calculated in step 1.[3] After identifying outliers, count the total number of outliers across all groups to provide an additional overall index of the similarity of schools. (See the "Should identified matched schools be limited to those within a maximum distance from the target school?" section above for a description of how calipers might reduce or prevent outliers.)

Step 7. Repeat steps 2–6 for all matching methods and measures of interest

Repeat steps 2–6 across the different matching options being evaluated. This produces three indexes of similarity for each measure of interest. Step 4 identifies the average absolute difference of each target school from its matched set, which provides an overall index of the average difference between target schools and their matched schools. Step 5 identifies the average standard deviation across matched groups, which provides an overall index of the similarity of matched schools. Step 6 identifies the number of target schools that are outliers within their matched groups, which provides an additional index of similarity.

Step 8. Evaluate the matching options

After calculating the summary statistics for each matching option, evaluate the quality of matches produced by different matching options. In general, the best matching option have these three outcomes:

1. Matched schools that are very close to their target school on the mean level of the measures of interest used to evaluate similarity (the mean of absolute differences across groups of matched schools from step 4 is small[4]).

2. Matched schools are very close to one another (the average of the standard deviations of matched groups from step 5 is small).

3. The fewest outliers from step 6.

Large differences between target schools and their matched schools, large variation between matched schools, or a large number of outliers indicates a less successful matching option.

3. If the matched groups have a small number of schools (fewer than six)—as a result either of choice or of a matching method such as calipers that can limit the number of matches—then the square root of the average of the squared standard deviations across groups calculated in step 5 is likely a better alternative to the standard deviation of individual groups calculated in step 2, which can have little meaning when the group is small.

4. It might be easier to rank these values by calculating a standardized average of absolute difference as the average of absolute difference divided by the standard deviation of the absolute differences.

Example of evaluating the matching options

This section provides an example of evaluating matching options and offers two ways to judge which option is more appropriate. Two matching options are evaluated: one uses the Euclidean distance method, and the other uses the Mahalanobis distance method. Each option includes multiple variables in the calculation of the distance measures, but only four measures of interest are used to evaluate the two options: national school lunch program eligibility rate from the student background characteristics variable category, per capita income from the community demographic characteristics category, enrollment from the school characteristics category, and attendance rate from the student performance category. The variables used in matching were standardized prior to calculating the distance measures.

Completing steps 2–7 for each measure of interest for both matching options results in sets of six summary statistics (table 4). The matching options are compared on each index of similarity for each measure of interest. For the national school lunch program eligibility rate the average difference between target schools and their matched schools was smaller for the Euclidean distance method (D_p = 0.02) than for the Mahalanobis distance method (D_p = 0.05). Therefore, the Euclidean distance method is ranked first and the Mahalanobis second. For the national school lunch program eligibility rate the Mahalanobis distance method produced matched schools with more similarity (M_{SD} = 0.17) than did the Euclidean distance method (M_{SD} = 0.30), so the Mahalanobis distance method is ranked first for this

Table 4. Example of summary results for evaluating different matching options

	Index of similarity					
	Average absolute difference[a]		Group similarity[b]		Outliers[c]	
Measure of interest and matching option	Value	Rank	Value	Rank	Value	Rank
National school lunch program eligibility rate						
Euclidean distance method	D_p = 0.02	1	M_{SD} = 0.30	2	n = 121	1
Mahalanobis distance method	D_p = 0.05	2	M_{SD} = 0.17	1	n = 235	2
Per capita income						
Euclidean distance method	D_p = 0.30	2	M_{SD} = 0.14	1	n = 123	1
Mahalanobis distance method	D_p = 0.22	1	M_{SD} = 0.16	2	n = 323	2
Enrollment						
Euclidean distance method	D_p = 0.23	1	M_{SD} = 0.24	1	n = 325	2
Mahalanobis distance method	D_p = 0.31	2	M_{SD} = 0.35	2	n = 125	1
Attendance rate						
Euclidean distance method	D_p = 0.11	2	M_{SD} = 0.05	1	n = 27	1
Mahalanobis distance method	D_p = 0.09	1	M_{SD} = 0.08	2	n = 37	2

a. The average of the absolute differences between each target school and its matched schools.

b. The average of the standard deviations of matched schools from step 5.

c. The count of schools whose value on the measure of interest is greater than two standard deviations from the mean of the target school's group of matched schools.

Source: Authors' construction.

index of similarity. Finally, on the same measure of interest, the Euclidean distance method produced fewer outliers (*n* = 121) than the Mahalanobis distance measure did (*n* = 235). Thus, the Euclidean distance method was ranked first for outliers. This process of ranking matching options is continued across the three indexes of similarity for the three remaining measures of interest (see table 4).

After ranking the matching options for each measure of interest and index of similarity, education agencies should consider two alternatives for choosing which matching option is best: evaluate the quality of the matches based on the relative importance of the measures of interest and the relative importance of the three indexes of similarity, or create a summary of the quality of matches that equally weights all measures of interest and all indexes of similarity.

The first alternative is to evaluate the quality of the matches based on the relative importance of the measures of interest and the relative importance of the three indexes of similarity. Because education agencies likely have different goals of identifying similar schools, each agency should choose a matching option that results in the identification of schools that are the most similar on the measures of interest and indexes of similarity that are the most relevant to the agency's context and needs. Some agencies might prioritize certain measures on which they want schools to be similar, such as demographic characteristics or performance. Other agencies might choose a matching option that minimizes the difference between a target school and its matched schools or a matching option that reduces the number of outliers.

If in the example displayed in table 4, the national school lunch program eligibility rate is the most important measure of interest and similarity among matched schools is the most important index, an education agency would choose the Mahalanobis distance method because it is ranked first on this index of similarity, even though Euclidean distance is ranked first on the other two indexes of similarity for the national school lunch program eligibility rate.

A second alternative is to create a summary of the quality of matches that equally weights all measures of interest and all indexes of similarity. This alternative summarizes which matching option performs better in general. Using the rankings of similarity described above, an education agency counts the number of times each matching option is ranked first. The matching option with the highest number of first rankings is the one that produces the most similar schools across all measures of interest and all indexes of similarity.

In the example in table 4 the Euclidean distance method is ranked first eight times, whereas the Mahalanobis distance method is ranked first four times. Thus, the Euclidean distance method produces the best overall quality of matches across the four measures of interest and three indexes of similarity.

Actions that could be informed by identifying similar schools

This guide offers considerations for deciding which variables to include and which distance method to use for identifying similar schools. The guide also describes a process for evaluating the quality of matches that result from different decisions (for example, including different variables or using different methods to calculate the distance measure). The guide was developed based on the approach that the NDE selected for identifying similar schools in Nebraska.

Education agencies interested in identifying similar schools might feel overwhelmed by the task and struggle to make informed decisions about what variables to include and which distance method to use. This guide can help agencies manage the process of making those decisions.

Education agencies might use the approach described here for identifying similar schools for a variety of purposes. For example, they might want to provide the most accurate group of similar schools to help education leaders better understand and interpret how their schools are performing. Agencies might also want to create opportunities for similar schools to connect and share practices with one another.

In addition to providing target schools with sets of matched schools to help them gauge school improvement efforts and student performance, education agencies could use the identified groups of similar schools to tailor supports to schools. For example, an education agency might identify a group of similar schools that are all struggling with the same issue, such as low high school graduation rates. The agency could then provide support that is customized to the student populations or other characteristics of the similar schools.

Finally, the approach for identifying similar schools described in this guide does not provide clear information about the education practices used by schools in the groups of similar schools identified for a target school or information about why some schools perform better than others on academic outcomes. Moreover, schools are likely to vary in a number of ways that are not captured in federal or state data systems, such as the level of parent involvement and the effectiveness of school leaders and teachers. Without access to reliable sources of information on such factors, they cannot be included in the approach for identifying similar schools.

Appendix A. Other example approaches for identifying similar schools

Research has frequently demonstrated associations between the demographic characteristics of a school population, such as the percentage of students living in poverty, and school performance (Hegedus, 2018). Consequently, researchers have argued that, to be fair, school leaders should consider the student population of a school when assessing its performance on academic outcomes (Darling-Hammond & Ascher, 1991; Hegedus, 2018; Salganik, 1994).

Many state education agencies provide both demographic information and student test scores in school report cards so that academic outcomes can be interpreted in light of the demographic characteristics of student populations. School and district leaders and community members can use this information to determine how schools perform in comparison with other schools with similar populations. However, making such determinations can be difficult because of the need to sift through multiple characteristics to identify similar schools for comparison (Salganik, 1994).

To lessen this burden, some state and local education agencies have identified and reported information about groups of similar schools.

New Mexico

The New Mexico Public Education Department (NMPED, 2017) also ranks schools relative to groups of similar schools. NMPED's ranking does not factor into a school's accountability rating. Rather, rankings are provided for informational purposes to foster collaboration between leaders at schools that struggle in a particular area (such as low math achievement) and leaders at similar schools that excel in that area. Such collaboration exposes leaders to possible strategies for improving outcomes.

To identify similar schools, NMPED uses latent class analysis. NMPED first categorizes schools as elementary, middle, or high and then uses these classifications to form "clusters" within each of these categories. The variables used to define the clusters include percentage of English learner students, percentage of students with disabilities, student mobility, economic disadvantage (based on median household income), and percentage of racial/ethnic minority students. Using this process in 2017, NMPED identified six elementary school clusters, three middle school clusters, and five high school clusters. Additionally, NMPED included a single cluster for all supplemental accountability model schools.

New York City

To evaluate charter schools for reauthorization, the New York City Department of Education (NYCDOE) considers how the charter schools are performing relative to a group of similar schools (Office of School Design and Charter Partnerships, n.d.). NYCDOE creates comparison groups that can be used to estimate how students in a charter school might have performed if they had attended other city schools and to evaluate how the students actually performed relative to a larger, hypothetical group composed of very similar students (New York City Department of Education, 2017).

For each main student NYCDOE first identifies a group of students who are an exact match on grade level, English learner status, disability status, and homelessness or eligibility for public assistance. From this group of exact matches NYCDOE uses primary factors (English language arts and math assessment scores) and secondary factors (the school's economic need and its percentages of students with disabilities and English learner students) to find the 50 students most similar to the main student. The primary factors are weighted more heavily than the secondary factors. For example, NYCDOE uses English language arts and math performance in a first round of comparison. If this results in a group from which 50 most similar students cannot be identified, then NYCDOE uses secondary factors as a second round of comparison to refine the list. From this list, NYCDOE creates a larger comparison group of all similar students for each student in the school. For example, if a school has 100 students, each of those 100 students will have 50 students from across the city identified as most similar. This comparison group will contain all of these most similar students. The comparison group has an upper bound size of 5,000, as matching is done with replacement. The school's performance is then compared with the performance of the students in the comparison group.

Oregon

Like the education agencies in Texas and New Mexico, the Oregon Department of Education (ODE, n.d.) starts the process of identifying similar schools by categorizing schools as elementary, middle, high, or combined schools (a combination of high school grades and grade 7 or lower). ODE then uses principal component analysis and four student variables (percentage of students eligible for the national school lunch program, percentage of students identified as English learner students at any point, percentage of students belonging to an underserved racial/ethnic group, and percentage of students who are mobile within a school year) to derive two component scores. The two component scores are determined for each school, and then ODE calculates the Euclidean distance between each pair of schools in the state. Because each of the four variables is a percentage, no standardizing is needed. For each school ODE applies a size filter (or caliper) to remove schools with student enrollment that is substantially different (larger or smaller) from that of the target school. As a final step, to limit the size of the comparison groups, ODE includes a maximum of 20 comparators identified as the closest in Euclidean distance to the target school.

Texas

In the Texas Accountability System the Texas Education Agency (TEA) identifies a group of 40 comparison schools for every school across the state (Division of Performance Reporting, 2017). The top-performing schools in each group earn a distinction designation.

To determine a “campus comparison group,” TEA first categorizes each campus by school type—elementary, elementary/secondary (K–12), middle school, or high school. TEA then groups each campus with 40 other campuses, from anywhere in Texas, identified as most similar using measures of grade levels served, size, percentage of students who are economically disadvantaged, student mobility rate, percentage of English learner students, percentage of students receiving special education services, and percentage of students enrolled in an Early College High School program.

To calculate what TEA calls the “linear distance” between schools, it uses the Euclidean distance method. As a first step, to account for the different scales of the included variables, TEA computes uniform linear values by standardizing each variable to a range of 0–100. TEA also uses matching with replacement so that a school might be included in an unlimited number of comparison groups.

United Kingdom

The U.K. Department for Education (DfE, 2018) begins its identification of “most similar schools” by grouping schools by school type (for example, academy, independent school, free school) and level (for example, primary, middle, secondary). The second grouping is by percentage of students with an education, health, and care plan and the percentage of students who have been eligible for free school meals at any point in the last six years. Using these two statistics, DfE calculates the Euclidean distance between each pair of schools. Because both of these statistics are measured as percentages, no standardization is required. From these two statistics DfE identifies the 49 most similar schools to each target school, forming a group of 50 most similar schools. In identifying inputs for the School Efficiency Metric, DfE seeks to include only measures within a school’s control and considers the extent to which schools contribute to student achievement.

Appendix B. The Euclidean and Mahalanobis distance methods

The second decision point in the approach described in the main part of this guide focuses on selecting an appropriate method for calculating a distance measure. Two methods are presented: the Euclidean distance method and the Mahalanobis distance method. This appendix provides more details about these two methods as well as additional information about the advantages and disadvantages of each.

Euclidean distance in a multidimensional space is represented by the following formula:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_i - y_i)^2 + \ldots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

where d is the distance between any two schools, x and y, on each variable i for a total of n variables.

The Mahalanobis distance method was developed to calculate the distance between a point and a distribution. Mahalanobis distance is a multivariate equivalent of the Euclidean distance and measures how far a given point is from the mean of the distribution. The distance is zero if the point is located exactly at the mean of the distribution.

The Mahalanobis distance method can also be used to calculate the distance between two points in multivariate space. For finding the distance between two points, x_A and x_B, Mahalanobis distance can be represented by the following formula:

$$MD = \sqrt{(x_B - x_A)^T \cdot C^{-1} \cdot (x_B - x_A)}$$

where $(x_B - x_A)$ is the matrix of differences between the points, T indicates a transposed matrix, and C^{-1} is the inverse variance–covariance matrix.

Advantages and disadvantages of the two distance methods

If the variables selected to identify similar schools are uncorrelated (for example, student achievement and school grade-level configuration), the Euclidean distance method can be an appropriate choice. But one disadvantage of the Euclidean distance method is that it ignores covariance (correlation) among variables. When Euclidean distance is applied to correlated data, information contained in the correlation will be duplicated so that variance along the correlated variables will be counted for both variables in calculating the distance.

For example, suppose there are hypothetical unobserved traits believed to drive outcomes. Additionally, each of the unobserved traits is expected to have a similar influence on the outcomes. Ideally, if it is known that two correlated variables are capturing two distinct traits and each trait deserves equal weight, then each variable should probably be given equal weight. However, because the traits are unobserved, evidence that two variables are correlated is often taken as evidence that they reflect a single underlying trait, especially if the variables are conceptually similar. If they do represent a single underlying trait, they probably should be given less weight than if they had a low correlation. As a result, including both variables when calculating the Euclidean distance might result in loss of meaning for the distance calculated (De Maesschalck et al., 2000; McCune & Grace, 2002), unless there is a desire for this information to be explicitly duplicated.

Another limitation of Euclidean distance is that, for included variables, the scales of the variables can be somewhat arbitrary and yet the scale affects how much the variable matters in the distance calculation relative to other variables. For example, when eligibility for the national school lunch program and English learner status are included as variables, the scales for these variables could be a percentage for both, a fraction for both, or a percentage for one and a fraction for the other. If a variable is measured as a percentage (0–100), it will matter far more in calculating the distance than if it is measured as a fraction (0–1). Similarly, financial variables could be reported in dollars, thousands of dollars, or millions of dollars. Those variables will have far more influence on the distance if they are measured in smaller units, like tens of dollars, than if they are measured in larger units, like millions of dollars (see the example in table 2 in the main text). As explained in the main text, the Mahalanobis distance method addresses both of these limitations.

The Euclidean distance method might have two advantages over the Mahalanobis distance method. The first is that it is computationally easier. In certain instances when calculating Mahalanobis distance, computation of the variance–covariance matrix can create challenges. One challenge occurs when a dataset includes a large number of variables, which can result in a great deal of redundant or correlated information. In some cases this multicollinearity in the data can lead to a singular or nearly singular variance–covariance matrix that cannot be inverted. Consequently, a pseudoinverse needs to be estimated, or Mahalanobis distance cannot be calculated. When Mahalanobis distance is used to calculate the distance measure for pairs of schools, singular covariance matrices are unlikely to be a problem. If a problem does arise, however, it might be easily remedied by dropping the offending variable (for example, one racial/ethnic category when all racial/ethnic categories are included). A second challenge is that the number of elements in the dataset must be larger than the number of variables (De Maesschalck et al., 2000).

The second advantage that the Euclidean distance method might have over the Mahalanobis distance method is that it enables equal weighting of all variables, even if they are correlated. This might matter if, for example, variables such as gender and English learner status are correlated. In this case researchers might argue that those variables should be given equal weight relative to a third variable to which they are not correlated, such as eligibility for the national school lunch program. The equal weighting might be desirable, even though the two variables are correlated with each other, based on the theory that they are capturing distinct underlying traits and that all distinct traits should be given equal weight.

Calculating Mahalanobis distance: An example

The main part of this guide includes examples of how to calculate Euclidean distance. Calculating Mahalanobis distances is more complex because it involves creating the inverse variance–covariance matrix in order to complete the calculation. As an example of how to calculate Mahalanobis distance to determine the distance between two points (or schools), consider the following set of data showing values for three variables for 23 schools (table B1).

There are several software programs that will compute Mahalanobis distance from a set of data, but to understand the process, users need to understand several steps.

Table B1. Example measures for three variables from 23 schools

School	National school lunch program eligibility rate (percent)	Per capita income (dollars)	Enrollment (number of students)
1	32.36	44,116.28	364
2	63.43	16,216.35	4,975
3	34.36	17,269.44	363
4	90.25	3,245.37	3,310
5	17.08	43,911.04	5,194
6	19.03	41,703.16	6,266
7	2.51	45,275.85	50
8	12.59	636.15	2,016
9	17.23	43,151.70	8,142
10	50.39	24,628.29	8,653
11	42.67	30,126.47	4,708
12	38.48	32,153.90	9,023
13	71.60	11,894.73	6,606
14	5.93	45,310.11	1,361
15	83.81	10,282.32	4,212
16	69.59	16,943.63	656
17	30.62	33,851.24	7,470
18	27.60	38,508.23	3,245
19	2.49	47,028.34	5,116
20	79.52	2,214.04	772
School A	24.00	35,000.00	4,862
School B	50.00	25,000.00	4,384
School C	75.00	35,000.00	258

Source: Authors' construction.

Step 1: Calculate the variance for each independent variable

The first step is to calculate the variance for each independent variable because the variances will be needed to create the variance–covariance matrix. The VAR.S function in Microsoft Excel can be used to estimate the variance of a sample. For this sample the national school lunch program eligibility rate variance is 756.81, the per capita income variance is 234,147,178.20, and the enrollment variance is 8,302,944.02.

Step 2: Calculate all possible covariances

In this example there are three variables, so there are three possible covariances to calculate. Again, these values are needed to create the variance–covariance matrix in the next step. In Excel the COVARIANCE.S function can be used to calculate the covariance of a sample of data point pairs.

- Between national school lunch program eligibility rate and per capita income, the covariance is −296,800.07.
- Between per capita income and enrollment, the covariance is 6,029,782.42.
- Between national school lunch program eligibility rate and enrollment, the covariance is −6,300.54.

Step 3: Create the variance–covariance matrix

Because there are three variables in this dataset (national school lunch program eligibility rate, per capita income, and enrollment), the variance–covariance matrix will be a 3 × 3 matrix, which can be conceptualized as shown in figure B1.

Figure B1. Values used to create variance–covariance matrix

Variance of national school lunch program eligibility rate	Covariance between national school lunch program eligibility rate and per capita income	Covariance between national school lunch program eligibility rate and enrollment
Covariance between national school lunch program eligibility rate and per capita income	Variance of per capita income	Covariance between per capita income and enrollment
Covariance between national school lunch program eligibility rate and enrollment	Covariance between per capita income and enrollment	Variance of enrollment

Source: Authors’ construction.

Based on the variance and covariance values calculated in steps 3 and 4, the variance–covariance matrix for this data sample is:

$$C = \begin{bmatrix} 756.81 & -296800.07 & -6300.54 \\ -296800.07 & 234147178.20 & 6029782.42 \\ -6300.54 & 6029782.42 & 8302944.02 \end{bmatrix}$$

In Excel this matrix is represented in a 3 × 3 grid of nine cells, with each value in a different cell.

Step 4: Create the inverse of the variance–covariance matrix

The inverse of the variance–covariance matrix created in step 3 is needed to calculate the Mahalanobis distance. Creating the inverse of a matrix is a complex mathematical process that involves calculating the matrix of minors, turning that into the matrix of cofactors, transposing this to result in the adjugate, and finally multiplying by the determinant. Explaining this process is beyond the scope of this guide. When computed manually, these calculations often result in errors and so are better left to statistical software programs. In Excel the MINVERSE function produces the inverse of a matrix. The inverse variance–covariance matrix for this set of data is:

$$C^{-1} = \begin{bmatrix} 0.006 & 1.17E{-05} & 4.74E{-06} \\ 1.17E{-05} & 2.56E{-08} & 5.45E{-09} \\ 4.74E{-06} & 5.45E{-09} & 1.26E{-07} \end{bmatrix}$$

Step 5: Identify the target school

For this example school A is selected as the target school. Table B1 reveals that school A has a national school lunch program eligibility rate of 24.00 percent, a per capita income of $35,000.00, and an enrollment of 4,862.

Step 6: Calculate Mahalanobis distances

Recall that the formula for calculating Mahalanobis distance is:

$$MD = \sqrt{(x_B - x_A)^T \cdot C^{-1} \cdot (x_B - x_A)}$$

Calculating the Mahalanobis distances between school A (the target school) and the other schools in the sample (see table B1) requires having the matrix of differences between school A and each matched school. For example, school 1 has a national school lunch program eligibility rate of 32.36 percent, a per capita income of $44,116.28, and an enrollment of 364. Because there are three variables in this dataset, the matrix of differences for school A and school 1 will be a 3 × 1 matrix, with each row in the matrix containing the difference between the value of school A for a variable and the value of school 1 for the same variable:

$$\begin{bmatrix} 32.36 - 24.00 \\ 44116.28 - 35000 \\ 364 - 4{,}862 \end{bmatrix} = \begin{bmatrix} 8.36 \\ 9116.28 \\ -4498 \end{bmatrix}$$

These differences can be easily calculated using Excel.

The Mahalanobis distance between these two schools can now be calculated as:

$$MD_{A,1} = \sqrt{[\,8.36 \quad 9116.28 \quad -4498\,] \cdot \begin{bmatrix} 0.006 & 1.17E{-}05 & 4.74E{-}06 \\ 1.17E{-}05 & 2.56E{-}08 & 5.45E{-}09 \\ 4.74E{-}06 & 5.45E{-}09 & 1.26E{-}07 \end{bmatrix} \cdot \begin{bmatrix} 8.36 \\ 9116.28 \\ -4498 \end{bmatrix}}$$

Excel's MMULT function can be used to multiply matrices, as is necessary in the above calculation. Recall that matrices are multiplied by calculating the dot product between two matrices and that operations in math are completed from left to right. Thus, the MMULT function should be used first to multiply the first and second matrices and then to multiply the resultant matrix by the third matrix.

$$MD_{A,1} = \sqrt{[\,8.36 \quad 9116.28 \quad -4498\,] \cdot \begin{bmatrix} 0.006 & 1.17E{-}05 & 4.74E{-}06 \\ 1.17E{-}05 & 2.56E{-}08 & 5.45E{-}09 \\ 4.74E{-}06 & 5.45E{-}09 & 1.26E{-}07 \end{bmatrix} \cdot \begin{bmatrix} 8.36 \\ 9116.28 \\ -4498 \end{bmatrix}} =$$

$$\sqrt{[\,0.139223951 \quad 0.000306771 \quad -0.00047847\,] \cdot \begin{bmatrix} 8.36 \\ 9116.28 \\ -4498 \end{bmatrix}} = \sqrt{6.113} = 2.47.$$

So the Mahalanobis distance between school A and school 1 is 2.47.

Comparing distance measures

This guide presents two methods for calculating distance: Euclidean distance with nonstandardized or with standardized data and Mahalanobis distance. These distances can be calculated manually, as described above, but statistical software packages (for example, SPSS, SAS, and R) also contain methods for calculating both Euclidean and Mahalanobis distances with a given dataset.

Each method can result in identifying a different group of similar schools (table B2). For example, for the dataset in table B1 and with school A as the target school, school 17 is the closest to school A if the variables are not standardized prior to calculating the Euclidean distance, school 18 is the closest to school A if the variables are standardized prior to calculating the Euclidean distance, and school 19 is the closest to school A if the Mahalanobis distance method is used.

Table B2. Similar schools to target school A using different methods for calculating distance, by distance

School	Euclidean distance (nonstandardized)	School	Euclidean distance (standardized)	School	Mahalanobis distance
17	2849.80	18	0.13	19	0.79
18	3862.95	6	0.18	11	0.84
C	4604.28	2	0.21	B	0.89
11	4876.00	17	0.24	6	0.90
12	5041.27	9	0.25	18	0.95
6	6848.62	5	0.25	5	0.96
9	8786.85	10	0.27	17	1.06
5	8917.23	B	0.28	2	1.32
B	10011.45	1	0.30	9	1.46
1	10165.56	3	0.38	14	1.49
14	10888.33	8	0.41	10	1.78
10	11042.86	16	0.43	12	1.79
7	11346.76	13	0.50	13	1.81
19	12031.04	12	0.53	16	1.93
3	18292.45	11	0.55	7	1.97
16	18539.82	C	0.63	15	1.99
2	18784.03	14	0.66	4	2.17
13	23171.04	7	0.78	1	2.47
15	24726.30	19	0.78	20	2.49
4	31792.60	20	0.79	3	2.72
20	33040.13	15	0.95	C	4.15
8	34481.50	4	1.18	8	6.53

Source: Authors' construction.

Choosing between distance methods

The main part of this guide describes some factors that education agencies can consider in determining which method to use to calculate a distance measure for identifying similar schools. Because of the simplicity of calculating Euclidean distance, agencies often choose this method after first standardizing all the variables (Division of Performance Reporting, 2017; Oregon Department of Education, n.d.; U.K. Department for Education, 2018). However, agencies might consider the Mahalanobis distance method to be more appropriate when the variables being used are correlated (De Maesschalck et al., 2000; Wicklin, 2012).

Additionally, the Mahalanobis distance method standardizes all variables based on their variances. In contrast, the Euclidean distance method, which requires that variables be standardized first, allows education agencies to use other standardization options. Differences in how variables are standardized, how much variables are correlated, and how the covariance is handled during the calculation of the distance measure could lead to different results between the Euclidean distance method and the Mahalanobis distance method. In these instances the Mahalanobis distance method is typically recommended because it allows for the covariance between variables and arguably does a better job of standardizing variables.

References

Cook, T. D., Zhu, N., Klein, A., Starkey, P., & Thomas, J. (2020). How much bias results if a quasi-experimental design combines local comparison groups, a pretest outcome measure and other covariates?: A within study comparison of preschool effects. *Psychological Methods.* Advance online publication. http://dx.doi.org/10.1037/met0000260.

Darling-Hammond, L., & Ascher, C. (1991). *Creating accountability in big city school systems* (Urban Diversity Series No. 102). ERIC Clearinghouse on Urban Education, National Center for Restructuring Education, Schools, and Teaching. https://eric.ed.gov/?id=ED334339.

De Maesschalck, R., Jouan-Rimbaud, D., & Massart, D. L. (2000). The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, *50*(1), 1–18. https://doi.org/10.1016/S0169-7439(99)00047-7.

Division of Performance Reporting. (2017). *2017 accountability manual for Texas public school districts and campuses.* Texas Education Agency, Office of Academics. Retrieved January 27, 2020, from https://tea.texas.gov/2017accountabilitymanual.aspx.

Hegedus, A. (2018). *Evaluating the relationships between poverty and school performance*. NWEA. https://eric.ed.gov/?id=ED593828.

Luellen, J. K., Shadish, W. R., & Clark, M. H. (2005). Propensity scores: An introduction and experimental test. *Evaluation Review*, *29*(6), 530–558. https://doi.org/10.1177/0193841X05275596.

McCune, B., & Grace, J. B. (2002). *Analysis of ecological communities*. MJM Software Design.

New Mexico Public Education Department. (2017). *Similar schools – Informational only, not for school accountability.* https://aae.ped.state.nm.us/SchoolGradingLinks/1617/Technical%20Assistance%20for%20Educators/Similar%20Schools%20Guide%202017.docx.

New York City Department of Education. (2017). *School quality reports: Using "Comparison Group" results to better understand a school's performance*. Retrieved January 27, 2020, from https://infohub.nyced.org/reports/school-quality/school-quality-reports-and-resources.

Office of Data, Research and Evaluation. (2019). *Methodology to compare districts and schools: A technical report*. Nebraska Department of Education. https://cdn.education.ne.gov/wp-content/uploads/2019/02/Methodology-to-compare-similar-peer-school-districts.pdf.

Office of School Design and Charter Partnerships. (n.d.). *Accountability handbook for NYC DOE Chancellor-authorized charter schools: School year 2018–19*. New York City Department of Education. Retrieved January 27, 2020, from https://infohub.nyced.org/reports-and-policies/school-quality/charter-school-renewal-reports.

Oregon Department of Education. (n.d.). *Comparison schools*. https://www.oregon.gov/ode/schools-and-districts/reportcards/reportcards/Documents/rccomparisonschools_details_1718.pdf.

Salganik, L. H. (1994). Apples and apples: Comparing performance indicators for places with similar demographic characteristics. *Educational Evaluation and Policy Analysis*, *16*(2), 125–141. https://doi.org/10.3102%2F01623737016002125.

Schochet, P. Z. (2008). *Technical methods report: Guidelines for multiple testing in impact evaluations* (NCEE No. 2008–4018). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://eric.ed.gov/?id=ED501605.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21. https://dx.doi.org/10.1214%2F09-STS313.

U.K. Department for Education. (2018). *School Efficiency Metric: A technical note on the definition and calculation of school efficiency*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/739888/180903_Metric_Methodology_v2_.pdf.

United States Census Bureau. (2016). *Gini index*. Retrieved December 16, 2020, from https://www.census.gov/topics/income-poverty/income-inequality/about/metrics/gini-index.html.

Wicklin, R. (2012). *What is Mahalanobis distance?* Retrieved January 27, 2020, from https://blogs.sas.com/content/iml/2012/02/15/what-is-mahalanobis-distance.html.

Acknowledgments

This project would not have been possible without the leadership of Matt Hastings, former senior administrator for the Office of Data, Research and Evaluation at the Nebraska Department of Education, and the collaboration of staff at the department. The authors extend a special thanks to Hongwook Suh, Kunal Dash, and Justine Yeo. The authors also acknowledge the contributions of Charles Harding, Douglas Gagnon, Joshua Stewart, and Trudy Cherasaro at Marzano Research, as well as Benjamin M. Kelcey and Matthew Gaertner of the REL Central Technical Working Group.

REL 2021–096

July 2021

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-17-C-0005 by the Regional Educational Laboratory Central administered by Marzano Research. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Van Dine, D., Randel, B., & Klute, M. (2021). *A guide to identifying similar schools to support school improvement* (REL 2021–096). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Central. http://ies.ed.gov/ncee/edlabs.

This resource is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.