

Technical note

This site provides composition and segregation indices for public school children in the elementary grades (kindergarten through six). Children are classified into four major racial/ethnic groups, based on reports from state education agencies. These are non-Hispanic white, non-Hispanic black, Hispanic, and Asian. Data are provided for three time periods, 1968-70, 1989-1990 and 1999-2000.

Data Sources

For the school years 1989-90 and 1999-2000, the data were downloaded from the Common Core of Data of the National Center for Education Statistics. NCES provides data for individual schools. We have counted only the students in each school who are in elementary grades. NCES does not provide a tabulation by race/ethnicity for every grade, but only a school total. Therefore in those cases where a school included both elementary and non-elementary students (such as a K-12 school), we have had to estimate the racial composition of the elementary students. We did this by assuming that the elementary students have the same racial composition as the school total. Approximately 10 percent of the schools in the NCES database comprise both elementary and non-elementary grades.

For school years 1968-71, the data were provided from a study conducted by Dr. Franklin Wilson and Dr. Karl Taeuber of the University of Wisconsin, Madison. Their data files are at the level of school districts. We selected all schools identified as “elementary districts” in their database. Their files included counts of students by race/ethnicity, as well as segregation indices that they had computed from the original school-level files that are no longer available. The metropolitan level measures were estimated from these data (see below). In many cases the data for “elementary” grades incorporate upper grades – as long as the school included at least grade 6. Therefore the enrollments that we list for 1968-71 may not be comparable to the enrollments for later years. Through analyses of 1989-90 data, however, we determined that the percentage distribution across races and the values of segregation indices are little affected by whether grades 7-8 are included.

Free Lunch Data, 1999-2000

Many states did not provide data on eligibility for free lunches in 1989-90, and we believe that reports for some other states may have understated the number of eligible children in that year. Therefore our data on free lunches are limited to the 1999-2000 school year. For that year, data are not reported for the following states: Arizona, Illinois, Tennessee, and Washington. Further, we have not included in our calculations information for any metropolitan region in which less than 70% of schools reported valid free lunch data.

Computation of Indices

There are three types of segregation indices: Index of Dissimilarity (D), Isolation Indices (P*), and Exposure Indices (P*). The following formulas were used to generate segregation scores. The formulas given below are for white/black indices, but the same formulas may be used for all combinations of groups.

The dissimilarity index (D) measures whether one particular group is distributed across schools in the metropolitan area in the same way as another group. A high value indicates that the two groups tend to attend different schools. D ranges from 0 to 100. A value of 60 (or above) is considered very high. It means that 60% (or more) of the members of one group would need to switch elementary schools in order for the two groups to be equally distributed. Values of 40 or 50 are usually considered a moderate level of segregation, and values of 30 or below are considered to be fairly low.

Another measure of school segregation is a class of exposure indices (p*) that refer to the racial/ethnic composition of the elementary school attended by the average member of a given group. For example, the average Hispanic might attend elementary schools that are 40% Hispanic, 40% non-Hispanic white, 15% black, and 5% Asian. (Note that these various indices must add up to 100%.)

The “isolation index” is the group’s exposure to itself – for example, the percentage of black students in the school attended by the average black student. Other exposure indices measure the group’s exposure to other groups. Indices of exposure must range from 0 to 100, where a larger value means that the average group member attends elementary school with a higher percentage of children from the other group.

Index of Dissimilarity

$$D = \left(\frac{1}{2} \right) \sum \left| \frac{B_i}{B} - \frac{W_i}{W} \right|$$

Where:

B = the school district black population

B_i = the black population of school i

W = the school district white population

W_i = the white population of school i

Isolation Index

$$P^* = \sum \left(\frac{B_i}{B} \right) \left(\frac{B_i}{T_i} \right)$$

Where:

B = the school district black population

B_i = the black population of school i

T_i = the total population of school i

Exposure Indices

$$P^* = \sum \left(\frac{B_i}{B} \right) \left(\frac{W_i}{T_i} \right)$$

Where:

B = the school district black population

B_i = the black population of school i

W_i = the white population of school i

T_i = the total population of school i

Missing or Inconsistent Data

The historical data from the University of Wisconsin included missing values for some cases in some years. Where data for school year 1969-70 were available, we used it. If not, we used either 1968-69 or, in a few cases, 1970-71.

Because compliance with NCES reporting is voluntary for state education agencies, statewide gaps in the reporting of student racial composition occur on an annual basis. Student racial composition was not reported for Idaho for any year between 1989 and 1999. Therefore Idaho was omitted from our analysis. In 1989 schools in the following states did not report student racial composition: Georgia, Maine, Missouri, Montana, South Dakota, Virginia, and Wyoming. In 1999, schools in Tennessee did not report student racial composition. For these states we merged the student membership and racial composition data from the next year in which these variables were available. The table below shows the states that did not report racial composition for each time period, and the years in which data were extracted and added to the 1989-90 and 1999-2000 files.

1989-90	1999-2000
Montana, Wyoming (1990-91)	Tennessee (1998-99)
Missouri (1991-92)	
South Dakota, Virginia (1992-93)	
Georgia, Maine (1993-94)	

The historical data contain some internally inconsistent values. Certain adjustments were applied to them:

- a. When a school district has no elementary students for a group, the Index of Dissimilarity measures involving the group are treated as missing. The group's Isolation Index and the measures of exposure of this group to other groups are also treated as missing. Because there are no group members, exposure of other (non-missing) groups to this group is set to zero.
- b. When a school district has only one group, the Isolation Index for the group is 100 and the group's exposure to other groups is 0. All other indices are treated as missing.
- c. When the non-Hispanic white exposure to minority students is greater than the percentage of minority students (which is logically impossible), it is treated as missing.
- d. Non-Hispanic white exposure to minority students should be equal to the sum of whites' exposure to each minority group taken separately. When the calculated value provided in the file was more than two points lower or higher than the sum of exposures to individual minority groups, the value is treated as missing.
- e. When the non-Hispanic white Isolation is negative (another logically impossible value), it is treated as missing.
- f. For school districts with only one school, the only valid information is the number of students in each race/ethnic group. There can be no segregation between schools, so all Index of Dissimilarity measures are set to zero. Exposure is simply the percentage of each group in the total school population.
- g. The Exposure Index for minority students to non-Hispanic whites should be equal to a weighted average of each minority group's exposure to whites (where the weight is the size of that minority group's enrollment in the district). In some cases this expectation was violated. These were revised as follows:
 1. When calculated measure is greater than the percentage of non-Hispanic whites in the district (a logically impossible value), all exposure indices of minority children to whites are treated as missing.
 2. When only the weighted average or the original minority-to-white exposure is greater than the percentage of whites, the other measure is used.
 3. When neither the weighted average nor the original minority-to-white exposure is greater than the percentage of whites, but the difference between the two alternative measures is greater than two points, the weighted average is used.

Geographic Units of Analysis

This web site provides data for school districts and for metropolitan regions (more precisely: Metropolitan Statistical Areas, either MSAs or PMSAs, as defined by the U.S. Census Bureau in the year 2000).

Where we could establish that school districts had consolidated or fragmented over time, based on the geographic location and names of districts, we have recombined schools in 1989-90 and 1999-2000 to create equivalent units. School districts from the 1968-71 data set were linked to those from NCES using the following identifiers: the district NCES code, the zip code of the district office, and the name of the district.

In reporting metropolitan-level results, it is important to use a constant geography for the boundaries of metropolitan regions. We used the 2000 definitions for the metropolitan region, for the central city or central cities as defined by the Census Bureau, and for the suburban remainder of the metro. We matched individual schools to metro areas based on a public use file from the Census Bureau containing codes for the nation's 331 MSAs and corresponding State FIPS and mailing zip codes. Because the NCES CCD provides the State FIPS and mailing zip codes for each public school, we were able to match the schools to the MSA file. We then aggregated the school data to the MSA level.

A central city/suburban code provided by NCES was used to determine whether a school was in the city or suburban portion of the metropolitan region. Data were aggregated on this basis to create city and suburban indices in 1989-90 and 1999-2000.

Calculating segregation indices for metropolitan regions in 1968-71 required additional computation. The original data set provided information for school districts, not individual schools. Therefore data were aggregated to metropolitan regions for entire school districts. Because many districts are known to cross over the city-suburban boundary, it was not possible to calculate separate city and suburban indices for 1968-71.

Exposure Indices and Isolation Indices are simply weighted averages of group members' exposure to other groups. Therefore the school district-level measures in the original file can be aggregated precisely to metropolitan regions. One difficulty is how to handle missing data. Where data are available for most school districts in a metropolitan region, can a reliable exposure index at the metro level be calculated if data are missing for some school districts in that region? We have taken a conservative approach to missing data. If data are available for exposure indices of a specific group in districts where at least 90% of the region's group members are enrolled, we have calculated and provided the indices for that group. The most common missing values are indices of exposure of blacks, Hispanics, and Asians to whites in 1968-71.

Dissimilarity Indices can be calculated exactly only from the original school-level data. Therefore, the metropolitan Dissimilarity Indices for 1968-71 provided here are estimates based on simulations of data for individual schools.

These estimates were developed through models in which every district was assumed to have only two schools – a school in which one racial/ethnic group is over-represented, and a school in which it is under-represented. This is a reasonable assumption, because calculation of the Index of Dissimilarity for any district could be conducted in two ways that lead to the same result:

- First, as shown in the formula above, one can compute for every school the degree to which one group is over-represented or under-represented compared to another group. Then the absolute values of these amounts are summed and divided by two.
- Second, one could combine the enrollments for all schools in which a group is over-represented, and compute the degree to which the group is over-represented in this aggregate. One could then combine the enrollments for all schools in which a group is under-represented, and compute the degree to which the group is under-represented in this aggregate. The absolute values of these (two) amounts should be the same and equal to the Index of Dissimilarity as calculated in the more usual way.

Our approach is to estimate, for every school district, how many students of group A and group B are enrolled in each of the district's simulated two schools. We then compute metropolitan-level segregation indices from the simulated school-level data.

The simulation is based on two pieces of information. One is the total number of each group in the district. In the example below for blacks and whites these are expressed as $black_T$ and $white_T$. The other is the value of the Index of Dissimilarity (D) for the two groups in this district. In fact, there are multiple solutions to this simulation that would be consistent with this information because the size of each school is not fixed. We use an estimate that is approximately midway among these solutions, in which we assume that the first school includes at least half of the total district black population, plus an additional number that is larger in proportion to the extent of segregation in the district. We assume that the first school includes no more than half of the total district white population, less an additional number that is larger in proportion to the extent of segregation in the district:

$$black_1 = (.5 + (D/200)) * black_T$$

$$white_1 = (.5 - (D/200)) * white_T$$

The numbers of black and white students in the simulated second school are simply the remainders.

We tested the reliability of this estimate using real data from 1989-90 and 1999-2000. The average “simulated” value of D across all metropolitan regions was equal to the actual value of D. In most cases it was within one point of the actual D, and in very few cases was it off by more than three points. This reliability stems from partly from the fact that an important component of metropolitan segregation is segregation between districts. This component is accurately captured in the district-level data. Variations in how each group is allocated between two schools – as long as they preserve the district’s dissimilarity index – have little impact on the calculation of the metropolitan value.