Assessing Multiethnic School Segregation:<br>Measurement and Interpretation<br>Ross E. Mitchell<br>University of Redlands<br>Douglas E. Mitchell<br>University of California, Riverside

April 30, 2010

Paper presented at the Division L - Educational Policy and Politics / Section 2: Legal, Judicial, and Intergovernmental Issues paper session, "Courts and Equal Educational Opportunities," of the American Educational Research Association annual meeting, Denver, Colorado.

## Assessing Multiethnic School Segregation:

## Measurement and Interpretation

This paper examines the measurement and interpretation of multiethnic segregation in public schools. Following a brief review of the history and theoretical importance of alternative attempts to measure segregation in multiethnic environments, we use both Monte Carlo simulations and a substantial body of empirical data from the nation's $13^{\text {th }}$ largest metropolitan region to estimate the probability density function for examining the confidence bounds of the most important indices of segregation and to give substantive meaning to changes that have occurred in this region over the last four decades. Since housing and transportation constraints make it virtually impossible to produce perfect desegregation in any district with two or more schools we give special attention to the likely values for relevant measures of segregation when imposing reasonable criteria for successful desegregation (such as requiring that the ethnic composition of all schools deviate by no more than $15 \%$ from the ethnic composition of the district within which they are located). Our empirical data cover 551 schools in 53 school districts over the four decades from 1968 through 2007.

The analysis presented in this paper has three specific objectives:

1. Summarizing alternative approaches to measuring ethnic segregation to identify approaches with the most robust statistical properties while recognizing that other indices have more intuitively meaningful and legally defensible measure criteria for pursuing successful multiethnic desegregation.
2. Creating a Monte Carlo sampling program documenting the probability distribution and the range of values likely to occur in school districts meeting generally accepted desegregation standards.
3. Interpreting field data from a large metropolitan region showing that, despite major improvements in overall school desegregation, school districts have persistently operated outside the reasonable and legal guidelines covering multiethnic desegregation.

## Recent Work Describing and Documenting Segregation

Chief among those who have assessed and monitored whether levels of school segregation are declining, advancing, or chronic are Gary Orfield (e.g., Frankenberg, Lee, \& Orfield, 2003; Orfield, 1983; Orfield \& Lee, 2007; Orfield \& Yun, 1999) and Charles Clotfelter (e.g., Clotfelter, 1979, 1999, 2001, 2004; Clotfelter, Vigdor, \& Ladd, 2006), though many others have contributed as well (recent work includes, e.g., Logan \& Oakley, 2004; Logan, Oakley, \& Stowell, 2006, 2008; Mitchell, Batie, \& Mitchell, 2010; Reardon \& Yun, 2001, 2005; Reardon, Yun, \& Eitle, 2000). The summary statement to be derived from these studies, using various segregation measures, is: Substantial school desegregation occurred prior to or during the 1970s and was largely maintained until about 1990, after which time school populations became more segregated, though at levels substantially below those of the 1960s. Orfield and Clotfelter have monitored changing racial and ethnic composition of U.S. schools and they describe school segregation in two ways: whether schools are racially balanced relative to district-wide representation of groups, and the degree to which students of different racial or ethnic groups are exposed to or isolated from one another. As will be discussed further below, these two perspectives describe the major legal and social scientific constructs that have framed measures of school segregation. Moreover, as discussed by Logan, Oakley, and Stowell (2006) and Clotfelter, Vigdor, and Ladd (2006), careful attention to both balance and exposure is critical to the interpretation of changes in school composition and segregation levels.

Our present interest is in assessing levels of segregation, not with probing the causes or consequences of the changes that have been occurring. When authors or agencies draw attention to school segregation, it is important to know how levels of segregation are defined and compared.

Whether segregation is an outcome of some administrative, political, or social process, or is an indicator of the environment or context in which schooling takes place, or is an input for an achievement production function or other explanatory model of educational or social outcomes, there must be a definition and measure of segregation that enters into the evaluation. And though this point has been made by various authors for more than half of a century (e.g., Duncan \& Duncan, 1955; Reardon, 2006; Stearns \& Logan, 1986; Zoloth, 1976), we emphasize here that fully characterizing the degree or extent of segregation is quite complex, both conceptually and mathematically, making it important to choose a measure or measures consistent with that complexity.

As discussed by Reardon and co-workers (e.g., Reardon, 2006; Reardon \& Firebaugh, 2002; Reardon \& O'Sullivan, 2004; Reardon, Yun, \& Eitle, 2000), historically persistent measures of school segregation may no longer be appropriate to monitoring the sprawling multiethnic educational landscape of America today. Newer measures have moved beyond the earlier emphasis on two-group comparisons. It is no longer appropriate to limit attention to dichotomous comparisons such as whites vs. non-whites or whites vs. blacks (or some other identified group without considering all other groups). As these authors demonstrate, information theory has provided a model for describing segregation that captures key conceptual and measurement features necessary for a robust index. The information theory formalism based on early work by Theil (1972) lends itself especially well to multigroup measures. Multigroup generalizations for several other popular measures (e.g., the dissimilarity, Gini, and normalized exposure indices) have been developed and their suitability for various applications has been assessed (e.g., Frankel \& Volij, 2009; Mele, 2009; Reardon \& Firebaugh, 2002; Reardon \& O’Sullivan, 2004). To support administrative and legal decision making related to school desegregation, however, we note that there are important problems associated with interpreting clearly what information entropy describes (e.g., Goodman \& Kruskal, 1959, p. 147) and usefully presenting this interpretation to lay audiences (e.g., Dziuban \& Esler, 1983, p. 120).

## Problems of Description and Measurement

As research scholars wrestle with the knotty problems associated with moving from measurement of simple two-group ethnic segregation to assessing the extent to which schools with substantial enrollments of more than two ethnic groups are meaningfully desegregated, we find that substantive, common sense, interpretations of exactly what is being measured are often lost. It is becoming clear, however, that the available measures of school segregation are based on distinctively different conceptions of what aspects of school enrollment should be considered when assessing the extent to which a school or district has segregated ethnic groups. In the paragraphs that follow we review the most frequently used multiethnic indices, giving specific attention to how their conceptions of school segregation differ.

## Descriptions of Segregation Are Based on Diverse Definitions

What defines segregation such that we can measure its degree? Massey and Denton (1988) provide a vocabulary for discussing the dimensions of segregation. When undertaking to promote school desegregation, federal mandates initially focused on assessments of exposure in order to eliminate the isolation of racial and ethnic groups that found themselves wholly or substantially in separate schools (i.e., the one-race schools of de jure segregation; e.g., see Brown v. Board of Education, 1954, 1955; Dye, 1968; Edelman, 1973; Green v. Board of Education, 1968; for Hispanics as the one separated race, see Mendez v. Westminster, 1946; Valencia, Menchaca, \& Donato, 2002; Wollenberg, 1974; for predominantly non-white schools as one-race schools, see Keyes v. School District No. 1, 1973). Very quickly thereafter, however, assessment measures of evenness became the focus of attention as the U.S. Supreme Court sought to "eliminate from the public schools all vestiges of state-imposed segregation that was held violative of equal protection guarantees by Brown v. Board of Education..." (Swann v. Charlotte-Mecklenburg, 1971, p. 2; also see, e.g., Rossell, 1985, p. 218; Rossell \& Armor, 1996, pp. 279, 290; Welner, 2006, p. 349). Though "the constitutional command to desegregate schools does
not mean that every school in every community must always reflect the racial composition of the school system as a whole, [the court affirms that a racial balance target defined by mathematical ratios may serve as] a starting point in the process of shaping a remedy" (Swann v. Charlotte-Mecklenburg, 1971, pp. 24-25). That is, evenness or racial balance may not be fully attainable but it is an important way to think about segregation and its remedy.

Because segregated housing patterns may be among the vestiges of state-imposed school segregation, legal mandates were sometimes used to constrain racial imbalance in order to ameliorate this legacy. It was necessary to overcome neighborhood school attendance zones where they were reinforcing the geography of segregation. That is, it became important to analyze the spatial dimensions of racial or ethnic residential concentration in sometimes tightly contained geographic areas, the clustering of multiple racially identifiable neighborhoods into large contiguous zones or ghettos, and racial or ethnic group centrality within large cities where particular groups occupy the city center. Because social geography is as likely to be a consequence as a cause of school segregation, the spatial dimensions of concentration, clustering and centrality are generally seen as factors complicating the process of desegregation, but are generally not seen as important measures for monitoring acceptable school segregation levels. The geographic boundaries of school districts have been, and remain, the most important spatial dimension limiting remedies (e.g., see Bischoff, 2008; Milliken v. Bradley, 1974, 1977). Research on student transfers and "white flight" has documented the importance of these spatial boundaries, but spatial analysis measures have never been an important part of desegregation supervision and monitoring.

In California, the state from which our Riverside-San Bernardino metropolitan area data are derived, school desegregation policy sought "to avoid and eliminate segregation of children on account of race or color" (Calif. Admin. Code, Title 5 § 2010, 1962). As articulated in Jackson v. Pasadena City School Dist. (1963), the aim was to counter both de jure and de facto segregation:

Residential segregation is in itself an evil which tends to frustrate the youth in the area and to cause antisocial attitudes and behavior. Where such segregation exists it is not enough for a school board to refrain from affirmative discriminatory conduct. The harmful influence on the children will be reflected and intensified in the classroom if school attendance is determined on a geographic basis without corrective measures. The right to an equal opportunity for education and the harmful consequences of segregation require that school boards take steps, insofar as reasonably feasible, to alleviate racial imbalance in schools regardless of its cause. (emphasis added, p. 881)

In other words, to the extent possible, the distribution of students among the schools should not be spatially dependent within a district's boundaries. Identifying the degree of school segregation is achieved simply by measuring the extent of racial or ethnic imbalance in the schools. The degree of school desegregation to be attained is defined primarily as an aspatial problem of increasing exposure or evenness; compliance is achieved by eliminating racial imbalances across schools within the district with limited regard to the distances between residential neighborhoods or the distances between balanced and imbalanced schools.

There are two important facets to this discussion: 1) in the years following the Brown decisions, school segregation was initially measured in terms of racial isolation and gradually moved to assessing racial balance, and 2) to be useful, a segregation index needs to be "easily applied to policy decisions in ways that yield numbers of pupils to be placed in specified schools" (Dziuban \& Esler, 1983, p. 120). We keep these two concerns in mind as we explore why Duncan and Duncan (1955), Stearns and Logan (1986), Orfield (e.g., Orfield \& Lee, 2007), Clotfelter (e.g., 2004), and others feel it necessary to calculate multiple segregation measures to describe the multiple dimensions of interest.

## The Limits and Merits of Theil's H

In an earlier study describing the historical evolution of school desegregation and housing integration in the multiethnic Riverside-San Bernardino metropolitan area of Southern California, we confronted these measurement and interpretation challenges (Mitchell, Batie, \& Mitchell, 2010; Mitchell, Mitchell, \& Batie, 2009). When three or more racial or ethnic groups must be considered simultaneously in a school or residential population, a compelling summary index or measure of segregation has to be not only conceptually clear but also have a mathematical formulation that distinguishes random fluctuations from substantial segregation and allows us to distinguish the relative contributions of inter- and intra-district problems of student distribution. Reardon and Firebaugh (2002) demonstrate the technical superiority of Theil's information index $(H)$ as a measure of segregation satisfying axioms of disproportionality, inequality, and decomposability, and we employed this approach in our previous analyses. Unfortunately, despite Reardon and Firebaugh's assertion to the contrary, Theil's H is not intuitively meaningful to most observers. Reardon and Firebaugh (2002) offer as a meaningful interpretation the following:

H can be interpreted as a normalized likelihood-ratio measure of association between two nominal variables indexing group and unit memberships, respectively. It can also be interpreted as one minus the ratio of the average within-unit population diversity to the diversity of the total population. (p. 42)

Regardless of its technical accuracy, this interpretation of Theil's H is beyond the comprehension of most policy makers and professional educators, and it provides little guidance regarding where segregation problems might be located or how they might be addressed. In non-technical parlance, Theil's H measures how much we can learn about students' ethnic identities by knowing which schools they attend. Thus, if district enrollment is 50 percent white and there is no segregation at all (i.e., Theil's $H$ is equal to zero) then we can do no better than make a $50 / 50$ guess about a student's ethnicity by
knowing their school. If Theil's H reaches 0.5, however, we can, on average, be 75 percent accurate about students' ethnicities by knowing their schools (i.e., our uncertainty is reduced by 50 percent).

The numerical value of Theil's H is straightforwardly calculated, and differences among the various indices for estimating the level of school segregation can be clearly seen by reviewing the formulas used to calculate them. The equations for each index calculation are given in each of the respective sections, in turn, but first some basic notation is defined. Frequencies are symbolized by $n$; proportions are symbolized by $\rho$; the number of racial or ethnic groups is symbolized by $c$ (for columns); the number of schools is symbolized by $r$ (for rows); and $i, j$, and $k$ are subscripts for summation across groups or schools (or for a specific partition of the racial/ethnic group by school enrollment table, e.g., $k$ might index the specific column with the white enrollment frequencies):

$$
\sum_{i=1}^{r} \sum_{j=1}^{c} n_{i j}=n_{++}=\mathrm{N} ; \sum_{j=1}^{c} n_{i j}=n_{i+} ; \sum_{i=1}^{r} n_{i j}=n_{+j} ; \frac{n_{i j}}{\mathrm{~N}}=\rho_{i j}
$$

That is, the number of students of a particular racial/ethnic group in a particular school (joint frequencies) is $n_{i j}$; total district enrollment is $n_{++}$or N ; individual school enrollment totals (row marginal frequencies) are $n_{i+}$; specific racial/ethnic group enrollment totals (column marginal frequencies) for a district are $n_{+j}$; and enrollment proportions are in reference to the total enrollment in the district:

$$
\sum_{i=1}^{r} \sum_{j=1}^{c} \rho_{i j}=\rho_{++}=1 ; \sum_{j=1}^{c} \rho_{i j}=\rho_{i+} ; \sum_{i=1}^{r} \rho_{i j}=\rho_{+j}
$$

For the present section on Theil's $H$ (equivalent to the Uncertainty Coefficient; see SAS Institute Inc., 2008, pp. 131-132; also Reardon \& Firebaugh, 2002), the following equations are provided for calculating its numerical value:

$$
H=\frac{E_{C}+E_{R}-E_{R C}}{E_{C}}
$$

$$
E_{C}=\sum_{j=1}^{c} \rho_{+j} \ln \left(\frac{1}{\rho_{+j}}\right) ; E_{R}=\sum_{i=1}^{r} \rho_{i+} \ln \left(\frac{1}{\rho_{i+}}\right) ; E_{R C}=\sum_{i=1}^{r} \sum_{j=1}^{c} \rho_{i j} \ln \left(\frac{1}{\rho_{i j}}\right)
$$

where the $C, R$, and $R C$ distinguish the information entropies $(E)$ for the racial/ethnic diversity of the district (column marginal proportions), the school size diversity of the district (row marginal proportions), and the school-level racial/ethnic diversity of the district (joint proportions), respectively. That is, $H$ is the proportion of the district's racial/ethnic diversity (column entropy, $E_{C}$ ) that is reflected in the racial/ethnic diversity of students among each of the schools across the district (ranging from $H=0$ when it is proportionately reflected in each school to $H=1$ when each school completely segregates a single racial/ethnic group within it and has no other groups represented-100\% reduction in uncertainty of the racial/ethnic identity of each student given knowledge of the school in which students are enrolled).

Once this proportional reduction in uncertainty interpretation of Theil's H is recognized it is relatively easy to see that policy makers and educators are not really interested in trying to predict student identities. They want to know two other important bits of information about the nature of segregation in the school system: a) how many students would need to be reassigned in order to produce an acceptable level of desegregation, and b) where are the students located that need to be moved and where are the schools to which they should be moved (or removed)?

Because Theil's H can be partitioned to identify which ethnic groups are the most segregated, or to determine the extent to which segregation is found within rather than between school districts, H can play a very prominent role in ascertaining where responsibility for action should be located even though it does not helpfully describe the actions needed. With its partitionability, however, H is unique in helping to identify the victims of segregation and allocate political and administrative responsibility for providing remedies.

## The Limits and Merits of the Dissimilarity Index

The Dissimilarity Index (D) lacks the mathematical elegance of Theil's H, but it has an important virtue that makes it important to policy makers and educators - it identifies precisely what proportion of a school district's population would have to change schools in order for all segregation to be removed. While both H and D reach zero at the same point - when each school's student composition mirrors precisely the ethnic composition of the district - H departs from zero in a non-linear way with respect to the number of children who would have to be relocated while $D$ estimates that number precisely.

The multigroup dissimilarity index can be calculated using the following equation, which is linear in the difference between the proportion of students from each racial/ethnic group in a given school and the proportion that would be required to have perfect proportionate representation of each racial/ethnic group in each school (D; see Reardon \& Firebaugh, 2002):

$$
D=\frac{1}{\sum_{j=1}^{c} \rho_{+j}\left(1-\rho_{+j}\right)} \cdot \frac{1}{2} \sum_{j=1}^{c} \sum_{i=1}^{r}\left|\left(\rho_{i j}-\rho_{i+} \rho_{+j}\right)\right|
$$

With the D statistic in hand, a policy maker has reasonable guidance about how much effort might be required to desegregate. Where the policy target is a reasonable level of desegregation (however reasonable might be defined) rather than complete desegregation, however, the D statistic shows a substantial weakness. If policy makers recognize that student mobility, housing segregation and transportation costs make perfect desegregation impossible to achieve, the calculation of $D$ does not ascertain whether desegregation actions are being directed to the most segregated schools within a district, and thus, at quite modest levels, would permit districts with several schools to be within overall desegregation guidelines while still having one or more schools that would be found to be severely isolated by the earlier measures of racial isolation.

The Dissimilarity Index also lacks H's capacity to directly compare inter- and intra-district segregation levels. Thus, while helpful to district planners in addressing within district segregation it is
not as useful as H in addressing broader governance concerns. D also lacks an easy way of determining whether specific ethnic groups are most subject to the ravages of segregation. Making it difficult in a multiethnic environment of determining which groups may be moving toward successful desegregation while others are becoming more sharply segregated.

Where longitudinal data are available, both D and H can be used to track changes in the overall level of within district segregation. H is a superior measure, however, because it can be used to track regional and inter-ethnic changes as well as overall levels of segregation. So far, no index is able to unerringly compare levels of school district segregation when the numbers of schools change from one measurement period to the next. ${ }^{1}$

## The Limits and Merits of the Gini Coefficient Index

The Gini Coefficient (G), named for its Italian creator, is the oldest of the evenness indices. It approaches the definition of ethnic unevenness in a still different way. For the Gini Coefficient, ethnic unevenness represents the extent to which members of each ethnic group tend to be disproportionately concentrated in some schools and not in others. When only two groups are being compared this is a reasonably powerful index because it measures the proportion total population of both groups that are not distributed equally among the schools. Reardon and Firebaugh provide a multiethnic generalization of $G$, but it is much harder to give its values an intuitively meaningful interpretation.

The multigroup Gini index can be calculated using the following equation ( $G$; see Reardon \& Firebaugh, 2002):

$$
G=\frac{1}{\sum_{j=1}^{c} \rho_{+j}\left(1-\rho_{+j}\right)} \cdot \frac{1}{2} \sum_{j=1}^{c} \sum_{i=1}^{r} \sum_{k=1}^{r}\left|\rho_{k+} \rho_{i j}-\rho_{i+} \rho_{k j}\right|
$$

Though the multi-group G now registers the tendency for group concentration, it no longer points to specific group(s) or schools as the source of an elevated index. In short, $G$ is not perfectly correlated
with either D or H , and, beyond its historical interest, does not seem to add much to our ability to describe school segregation or identify how to fix it.

## The Limits and Merits of the Ethnic Isolation Index

A fourth definition of segregation is implied in the Exposure Index (usually labeled $\mathrm{P}^{*}$ ). This index calculates the extent to which students of each ethnic group are attending schools where they are disproportionately likely to encounter members of their own ethnic group (isolation) or members of other ethnic groups (interaction). The intuitive meaning of this index has the virtue of simplicity-it is fairly easy to understand that a student whose schoolmates are drawn more from his/her own ethnic group than would be true if the school's enrollment matched district wide proportions for each ethnic group is to some degree being isolated from the other groups.

The non-white isolation index can be calculated from the following equation ( $P^{*}$; see Massey \& Denton, 1988):

$$
P^{*}=\frac{1}{\rho_{+k}} \sum_{i=1}^{r}\left(\rho_{i k} \cdot \frac{\rho_{i k}}{\rho_{i+}}\right)
$$

Here, $k$ is the subscript for the non-white group (i.e., the race/ethnicity by schools table is first consolidated into two columns by aggregating across all non-white racial/ethnic groups, leaving a column for white enrollments and a column for non-white enrollments).

While the level of this isolation is typically averaged across all schools to get a measure of district-wide isolation, this index can easily be calculated for each individual school and can therefore be used to identify schools of more substantial isolation. Unlike D, however, this index does not directly specify how many students need to be moved in order to overcome the isolation that exists. The name of this $P^{*}$ index is a bit problematic in most contemporary settings, because "isolation" levels are only modest and thus students are not really isolated, but surrounded by a mix of students that lowers exposure to other ethnic groups.

Supervision and Monitoring Indices: Isolation and Imbalance
While H, D, G and P* all view multiethnic segregation in terms of how evenly ethnic group students are distributed across all of the schools in a district (or other region), there are two segregation measures that look at individual schools and, using various criteria, ascertain whether a given school should be viewed as segregated.

## The Limits and Merits of the 90+ percent non-white school measure

Measurement of racial isolation began with the U.S. Civil Rights Commission promulgation of a measure based on the "per cent of total Negro elementary pupils in schools which are 90-100\% Negro" (Dye, 1968, p. 142). That is, near or complete racial isolation of African American children in particular schools is how to recognize segregation and its degree is determined by knowing the proportion of African American children so isolated. However, in multiethnic California, where segregation of Asian American and American Indian students had been legal until the 1947 repeal of state law following the Mendez v. Westminster decision (1946; see Wollenberg, 1974), which struck down illegal de jure segregation of Mexican American students, it is clear that the Civil Rights Commission's racial isolation index is too narrowly defined. The U.S. Supreme Court, in Keyes v. School District No. 1 (1973), redefined racial isolation in terms of non-white students who share similar discrimination experience regardless of historical bases (i.e., Mexican American and African American in the Denver case), but racial balance definitions for remedy had become more prevalent by the time of the Keyes decision.

For this paper, a racially isolated (non-white) schools (NWS) proportion can be calculated using the following equation:

$$
N W S=\frac{1}{r} \sum_{i=1}^{r} \delta_{i}^{*}
$$

The $\delta^{*}$ function is either one (1) or zero (0) based on the following criterion, where $k$ identifies the column for the white racial/ethnic group:

$$
\delta_{i}^{*}=\left\{\begin{array}{l}
1 ;\left(1-\frac{\rho_{i k}}{\rho_{i+}}\right) \geq 0.9 \\
0 ;\left(1-\frac{\rho_{i k}}{\rho_{i+}}\right)<0.9
\end{array}\right.
$$

Or, more simply, the NWS is the total number of schools that are racially isolated (non-white percentage of $90 \%$ or greater) divided by the total number of schools in the district. This differs from the U.S. Civil Rights Commission definition in that $\delta^{*}$ is neither weighted by the number of non-white students in each school nor scaled to their district population. However, the rank-order correlation between the two definitions is almost perfect, so this simpler proportion of isolated schools (NWS) is used here.

As will be seen in our review of ethnic enrollment data in a large metropolitan region discussed in a later section of this paper, the identification of 90 to $100 \%$ concentrations of non-white students remains a useful and powerful indicator of the persistence of racial isolation in the public schools. This index is no respecter of district boundaries and can identify all the schools in a district as isolated if that district is more than $90 \%$ non-white, even though measured by $\mathrm{H}, \mathrm{D}, \mathrm{G}$ and $\mathrm{P}^{*}$ the district produces a completely even distribution of the ethnic groups found within its service area. It is, of course, a political values question-not a measurement question-as to whether our governance systems wish to identify and re-structure ethnically isolated schools. If that remains a priority, then Ethnic Isolation measurement remains an important way to assess school segregation.

## The Limits and Merits of Ethnic Imbalance Measurement

Though a large number of different measures of segregation have been proposed, we conclude our review with one that measures Ethnic Imbalance by testing whether the ethnic composition of any school in a district deviates by more than some set percentage from the overall ethnic composition of the district. This measure is not only intuitively easy to appreciate, it directs attention squarely upon the ethnically imbalanced schools and specifies exactly what needs to be done to overcome the imbalances.

Criteria for labeling a school ethnically out-of-balance have varied from time to time and place to place. In California, for example, the State Board of Education adopted a 15\% rule in 1969 (Calif. Admin. Code, Title 5 § 14021(c)), which stated that a school was racially imbalanced if the enrollment of any racial or ethnic group at that school was not within plus or minus $15 \%$ of the district-wide percentage for that group (also see Hendrick, 1975, pp. 227, 230). Though this strict numerical rule, the 15\% rule (Calif. Ed. Code §§ 5002-5003, 1971), was eventually repealed by ballot proposition (see Santa Barbara Sch. Dist. v. Superior Court, 1975), it had been in place long enough to establish a clear sense of just how much racial imbalance is excessive (which complicated litigation in some cases; see, e.g., NAACP v. San Bernardino City Unified Sch. Dist., 1976). Similar bounds on racial imbalance as those originally adopted in California are found in many desegregation orders across the United States (some as tight as $10 \%$ and others as liberal as 20\%; see, e.g., Berger, 1984; Giles, 1977; Smith \& Mickelson, 2000; Rossell \& Armor, 1996; Welner, 2006).

The racially imbalanced schools (RIS) proportion can be calculated using the following equation:

$$
R I S=\frac{1}{r} \sum_{i=1}^{r} \delta_{i}
$$

The $\delta$ function is either one (1) or zero ( 0 ); the previously identified $15 \%$ Rule (or any other $\Delta$ ) criterion is the basis for decision (i.e., if obeyed then $\delta_{i}=0$, and if disobeyed then $\delta_{i}=1$, for school (row) $i$ of the $r$ schools in the district):

$$
\left.\begin{array}{c}
\left(\rho_{+j}-\Delta\right) \cdot \rho_{i+} \\
0
\end{array}\right\} \leq \rho_{i j} \leq\left\{\begin{array}{c}
\rho_{i+} \\
\left(\rho_{+j}+\Delta\right) \cdot \rho_{i+}
\end{array}\right.
$$

That is, for district ethnic group proportions of less than 0.15 or greater than 0.85 , the full $\Delta$ range of the $\pm 15 \%$ rule is truncated when either $0(0 \%)$ or $1(100 \%)$ is reached. Or, more simply, the RIS is the total number of schools that are racially imbalanced (violate the $15 \%$ rule) divided by the total number of schools in the district.

This measure recognizes segregation as gross over- or under-representation of one or more racial or ethnic groups in a school and degree of segregation is an all or nothing phenomenon (i.e., the school is either in compliance with the $15 \%$ rule or it is not). Clearly, there is no subtlety here, nor is there any sense of a continuous scale describing degree of segregation, only a count of racially imbalanced schools.

## Summarizing the Measurement Issue

As previously noted, social scientists have always had their own measures of segregation, many of which predate the aforementioned policy measures. Massey and Denton (1988) reviewed the literature and identified a score of measures, though only a few were widely used. The persistent and popular measures are the dissimilarity index (D), the Gini coefficient (G), and the exposure indices ( $\mathrm{P}^{*}$ ). A measure that has gained recent popularity for multiethnic applications, but was not seen much before the work of Reardon and Firebaugh (2002), is Theil's information index (H). The D, G, and H indices are all considered measures of evenness while the $P^{*}$ family of indices measure exposure. That is, D, G, and $H$ are different ways of defining how the observed racial or ethnic distribution among schools differs from what would be expected if all groups were distributed in proportion to the district-level total enrollment, while the $P^{*}$ family of indices are ways of describing the probability that individuals of one group encounter or are exposed to individuals of another group or themselves. The $P^{*}$ index is called an interaction index when measuring how much members of one group are exposed to members of other groups and an isolation index when measuring the degree to which members of one group are exposed to other members of their same group.

None of these measures tell the same story, however, because each is built upon a different model of segregation. For multigroup indices of evenness, segregation is understood in terms of the exchanges of students among schools; desegregation would result from a series of progressive exchanges that result in each school's racial/ethnic group distribution becoming more like the district-
wide proportion without changing the total enrollment in each school. The dissimilarity index is linear in exchanges and represents the proportion of individuals who have to be moved to achieve district-wide proportional representation of all groups across all schools. As a consequence of this proportional exchange interpretation, $D$ is the most sensible of the evenness indices among those who administer districts that must desegregate. Though D does not provide an answer to how many students have to be moved to particular schools—none of the evenness indices provide school-specific information-it does help with the aggregate planning issue of identifying what proportion of the district population will have to move to fully desegregate.

Theil's information index is non-linear in exchanges. The index changes more when there are exchanges of students among schools where racial/ethnic group populations are most imbalanced than for exchanges close to balance; it represents a diminishing marginal return on progressive exchanges toward desegregation. This makes sense when we realize that another way to think about Theil's H is that it measures the proportional reduction in error associated with guessing the racial/ethnic identity of a student given knowledge of the school the student attends. That is, when school systems are perfectly segregated, knowledge of which school the child attends removes all (100\%) of the uncertainty in student racial/ethnic identity when we know which school the child attends. However, as schools become desegregated, our uncertainty about the specific racial or ethnic group to which the student belongs increases rapidly. When districts are substantially but certainly not fully desegregated knowledge of the school a child attends provides very little reduction in our uncertainty about the child's race/ethnicity beyond knowledge of the overall district racial/ethnic composition.

The rank order of segregation among districts is readily and reliably obtained from H . The other evenness measures, though giving similar rank ordering of districts, do actually produce a different rank order in some instances. Moreover, there is no direct interpretation of the numerical value of the H index. What compels analysts to adopt H as their preferred index is that it is decomposable. We can
compare between district segregation to within district segregation across a county or metropolitan region. We can also compare whether segregation is largely between whites and all other racial/ethnic groups, for example, or if there is substantial segregation among non-white groups as well. None of the other indices permits such comparisons across both nested levels of school organization and racial/ethnic group membership.

When it comes to measures of exposure $\left(P^{*}\right)$, the interaction index does not generalize to multigroup contexts, but the isolation index does when normalized (i.e., as a weighted sum of variance ratios; see Reardon \& Firebaugh, 2002). However, our interest is in the dichotomous definition of isolation in terms of white vs. non-white students, which is consistent with the Keyes decision (1973) as well as the earliest notions of what defined segregation. Namely, it is the isolation of non-white students in schools separate from white children that continues to be a valid concern.

## Empirical and Simulated Data from the Nation's $13^{\text {th }}$ Largest Metro Area

We turn in this section from a review of theory to the application of theory to simulated and real examples of school segregation. Theory guides the design of two computational approaches for exploring the behavior of Theil's H: 1) a Monte Carlo simulation for estimating the probability density function and expected value for Theil's H when students are randomly assigned to schools within a district and 2) a systematic exploration of the hypersurface defined by all possible values of Theil's H in order to identify the maximum value H can attain when districts are meeting nominal desegregation criteria (for a similar approach to studying within-school segregation using an exposure index, see Conger, 2005). Only Theil's H is investigated in this computational study because of its putative status as the most versatile segregation measure, which means that developing an understanding of its values is critical to extending its application and accurate interpretation. Our Monte Carlo simulation is used to estimate the probability density function for H obtained by randomly assigning students to schools in a representative school district containing ten school sites and four different ethnic groups. By repeatedly
making random assignments that preserve the given ethnic composition of the school district and maintain fixed enrollments for each school in the district we identify the most likely index values for Theil's H. This was done under two different conditions. In the initial pass we imposed no constraint on the allowable level of school segregation within the district, thus determining what values of H are likely to occur if students are assigned on a purely random basis to district schools. In a second pass a probability density function was created to estimate expected values of H when a reasonable criterion for meaningful desegregation is imposed on the school district—we used the commonly imposed $15 \%$ rule, used by California during an era of strongly encouraged school desegregation. A second algorithm was developed to determine the maximum value that H could reach if all schools within the district were required to meet the $15 \%$ rule. This is done by maintaining a fixed ethnic group composition for the district and fixed enrollments for the district schools while repeatedly exchanging students among the schools and testing each exchange to see if it raised the district H value. If an exchange increased the H value, it was preserved and a new exchange process was initiated until no further exchanges raised the district H value. Because there are multiple local maxima on the H hypersurface, multiple starting points were tested to identify the global maximum.

Following the discussion of estimates of the expected and maximum H values for a school district with complete and constrained random assignment of students, we will turn attention to how the different segregation indices provide divergent interpretations of the history of school segregation/desegregation in the Riverside-San Bernardino metropolitan region. In this analysis we calculate the values for Theil's H, Dissimilarity (D), Gini (G), and non-white isolation ( $P^{*}$ ) index values, together with the proportions of racially imbalanced schools (RIS), and racially isolated (non-white) schools (NWS) for each school district with two or more schools in the two-county region for each of the available data years. The history of desegregation efforts is tracked using each of these indices to highlight differences among them.

## Monte Carlo Simulation

The approximation of a probability density function $(p d f)$ for Theil's H , based on a histogram of simulated values, was obtained by calculating 10,000 replications of a random allocation of students in the average district size and composition appearing in the Riverside-San Bernardino metro area over the last 40 years. On average, districts in this metro region had about 6,600 students composed of: 5 percent Asians \& Others, 9 percent African American (black), 42 percent Hispanic and 44 percent white students. On average, schools ranged in size from one containing about 7 percent of the district students to one serving about 13 percent. The Monte Carlo assignments of students preserved the ethnic and school composition totals throughout. The distribution generated was through random placement of students in schools one student at a time.

The race/ethnicity of the students being allocated and schools to which they were assigned were determined by random draw using a uniform random number generator. First, a vector of racial/ethnic group identifiers was created by randomly selecting a group member (identifier) in proportion to its share of the remaining district population (i.e., sampling without replacement) until the entire population was identified in the assignment queue vector. Second, the school to which the racial/ethnic group member was assigned was selected at random from all schools as long as the school had not reached its fixed occupancy in which case random selection of available schools was repeated until one with capacity was identified. The fixed marginal frequencies constrained the maximum allowable allocation to any particular combination of race/ethnicity and school. When random allocations reach the maximum allowable in any cell, a new random draw was taken until the entire student population had been allocated within the allowed marginal constraints. Once the allocation (two-way contingency) table had been filled, Theil's H was calculated for the joint distribution. Theil's H values for each replication were saved and then binned to obtain a $p d f$ histogram.

As noted above, two different frequency distributions of H were calculated-one using completely random assignment of students and the other imposing the $15 \%$ rule to develop a probability distribution of H values associated with random arrangements of students in schools required to meet this minimum standard of desegregation. The programming code used for this random assignment process contains a parameter establishing the required level of ethnic balance in the schools. Parameter values other than 1.0 (completely random assignment) and .15 (constraining balance to $\pm 15 \%$ were tested, but the basic lessons to be drawn from this simulation exercise are fully explained by the two density functions presented here. The entire simulation process was programmed in Visual Basic for Applications and plotted using Microsoft Excel 2007 (VBA code may be requested from the authors).

## Simulation Results

Figure 1 presents the two pdfs obtained from random assignment of students to the Riverside-

San Bernardino metro region's historical "average" district under two conditions: 1) Random distribution of enrollment with no desegregation constraints (the blue line) and 2) Random distribution of enrollment with the $15 \%$ desegregation constraint rule applied.

Though the two distributions are definitely not identical they do overlap significantly. We note that none of the 10,000 replications produced without imposing the $15 \%$ ethnic balance constraint resulted in a single school with an ethnic composition violating the $15 \%$ rule. That is, in this regional average school district, completely random allocation did not result in any racially imbalanced schools. When students are assigned to schools at random, the expected value for H is miniscule ( 0.0017 unconstrained; 0.0013 with $15 \%$ rule), indicating that random student assignment is entirely unlikely to produce significant ethnic segregation. Indeed, as discussed below, these expected values are lower than $99.3 \%$ of the 1,207 district-year observed values (which range from 0.0003 to 0.4247 ) found in the forty year historical record for the Riverside-San Bernardino metropolitan area.

In our second series of computations, calculating the maximum possible value that Theil's H can attain in this regional average district without violating the $15 \%$ rule for ethnic balance shows that this maximum value is more than 120 times as large as the average value produced by random allocation of students $\left(\mathrm{H}_{\mathrm{Max}}=0.163 ; \mathrm{H}_{\mathrm{Avg}}=0.0013\right)$. This maximum H value was greater than $95 \%$ of the empirically observed district-year values for the Riverside-San Bernardino public schools from 1968 to 2007. In short, it is possible, though not very likely, for school districts to have H values far in excess of those produced by random allocation of students without producing any substantially segregated schools.

Figure 1. Histogram of Monte Carlo simulated Theil's H values for the "average" public school district when constrained by the $15 \%$ rule and when unconstrained, Riverside and San Bernardino Counties, 1968-2007.


To summarize, the random allocation of 6,600 students among the 10 schools in a district of "average" composition, yields an overwhelming number of Theil's H values that represent very close to
perfect independence; i.e., very close to zero. Observed ethnic enrollments are clearly not the result of random assignment processes, and where circumstances are right an H value that is more than 120 times the randomly expected value can be produced without actually producing any ethnically imbalanced schools.

What looking at this single "average" district does not reveal is whether ethnically imbalanced schools would ever be observed at random when districts are far from the average size, number of schools, racial/ethnic composition, or relative size of enrollment in each school within the district. The probability of ethnically imbalanced schools generated at random is readily understood in terms of a chisquared distribution for two-way contingency tables. When school sizes are near the smallest observed in any district-year (here, this means 100 students), which necessarily makes district total enrollments small, the Monte Carlo simulation returns no more than a few percent of the replications with at least one ethnically imbalanced school—few violations of the $15 \%$ rule. This is simply a small sample size effect. Increasing the number of students enrolled in the smallest schools by fifty percent reduces this already small fraction of ethnically imbalanced schools by a factor of ten (i.e., to about 0.05\% or replications). That is, the probability of deviations from the most probable random distribution diminishes quickly with increasing sample size. In line with increases in the table degrees of freedom (here, in proportion to the number of schools since the number of racial/ethnic groups is constant), deviations from perfect proportionality are more likely when there are more schools in the district, which means that the proportion of ethnically imbalanced schools generated at random will go up a little when there are more schools in the district. The practical consequence of knowing all this is that ethnically imbalanced schools by random assignment would never be expected in urban school districts where school sizes are very large even though the number of schools is also high. Suburban and urban fringe districts in Southern California, which often have large elementary schools sizes as well, would be even less likely to have ethnically imbalanced schools due to random assignment.

We have also explored the extent to which the central tendencies and extreme values of Theil's H distributions are sensitive to whether a school district's ethnic composition varies, district school enrollments are substantially unequal in size, or the number of schools within the district changes. Because H is defined in relation to the racial/ethnic diversity of the district $\left(\mathrm{E}_{\mathrm{c}}\right)$, the less diverse is the distribution (more dominated by a single group) the higher is the expected (mean) value of H under random assignment. This is largely because small fluctuations in the total diversity ( $E_{\mathrm{RC}}$ ) are magnified by the small denominator ( $\mathrm{E}_{\mathrm{c}}$ ). Because increasing the number of schools necessarily increases both the school enrollment diversity ( $\mathrm{E}_{\mathrm{R}}$ ) and the total diversity ( $\mathrm{E}_{\mathrm{RC}}$ ), the expected value of H shows very little dependence on the number of schools. Over the range of realizable school size variation within districts (i.e., limited range of enrollment diversity, $\mathrm{E}_{\mathrm{R}}$ ), expected H values are effectively constant. However, when it comes to the extreme values, namely our interest in the maximum attainable H value, variations due to composition, enrollment, and size provide little information more important than what we have presented with the example of the "average" district. The maximum attainable value of Theil's H is two orders of magnitude greater than the expected value of H . And relevant to interpretation, the maximum attainable value is greater than the overwhelming majority of observed H values for districts that have ethnically imbalanced schools.

## Comparison of Simulated and Empirical Results for Selected District-Years

Review of some aspects of the historical dynamics of segregation and desegregation in our large metropolitan region will shed some additional light on the difficulties of interpreting Theil's H and the other prominent indices for measuring the degree of segregation present in a school district or region.

When it comes to making sense of current or historical circumstances, some specific cases may be helpful. All are ten-school districts, which holds constant the number of schools across which to distribute enrollment, but each has a distinctive racial/ethnic composition as well as some greater variability in the relative enrollment size at each elementary school throughout the district. The
racial/ethnic distribution for each selected district-year is provided in Table 1 under Percent Composition. The Desert Sands Unified School District (1989) is a majority Hispanic/Latino school district $\left(E_{C}=0.852\right)$. The Hesperia Unified School District (1992) is a majority white districts $\left(E_{C}=0.771\right)$. The Upland Unified School District (2001) is a plurality white school district having Hispanic students as its second largest group as well as an appreciable percentages of black students enrolled, and is the most diverse of the three ( $E_{C}=1.176$ ). The school enrollment distribution for each selected district-year is summarized under Percent Enrollment. Hesperia Unified-1992 has the greatest range of school sizes ( $\mathrm{E}_{\mathrm{R}}$ $=2.255)$ while Upland Unified-2001 has the narrowest range of school sizes $\left(E_{R}=2.293\right)$, though hardly different from Desert Sands Unified-1989 ( $\left.E_{R}=2.294\right)$. As can be seen from the row of observed Theil's H values, the district-years are listed from left to right in descending order of their Theil's H values. From the complete 10-school district empirically observed distribution of Theil's H values, Desert Sands Unified-1989 is at the $90^{\text {th }}$ percentile, Upland Unified-2001 is at the $75^{\text {th }}$ percentile, and Hesperia Unified-1992 is the lowest ( $2^{\text {nd }}$ percentile).

The unconstrained expected Theil's H values are quite small and significantly less than the observed values (i.e., observed values are not within the 95\% probability bounds, two-tailed, of the expected values). The $15 \%$ rule constrained Theil's H expected values are even smaller. In nearly all cases, the observed Theil's H values are noticeably less than the maximum possible value that may be obtained within the constraints of the $15 \%$ rule. The exception, Desert Sands Unified in 1989, is more segregated than would be possible for even the most extreme racial/ethnic distribution that still satisfies the $15 \%$ rule. An inspection of racial/ethnic distributions across schools for this district-year reveals that 3 of the 10 elementary schools are majority white in this majority Hispanic district while a fourth is a strong plurality white (49.6\%), a fifth is a racially isolated school (non-white enrollment $90.3 \%$ ), and two others are racially imbalanced (RIS $=0.7$, i.e., altogether 7 out of 10 schools violate the 15\% rule).

Table 1. District enrollment distribution descriptive statistics for selected ten-school district-years,

Riverside-San Bernardino metropolitan area, 1968-2007.

|  | District-Year |  |  |
| :---: | :---: | :---: | :---: |
|  | Desert Sands Unified1989 | Upland Unified2001 | Hesperia Unified1992 |
| District Size | 7,206 | 6,561 | 7,824 |
| Percent Composition |  |  |  |
| Asian \& Others | 1.51 | 7.71 | 1.42 |
| African American | 2.82 | 11.68 | 3.50 |
| Hispanic/Latino | 56.84 | 35.82 | 24.77 |
| White | 38.83 | 44.80 | 70.31 |
| Percent Enrollment |  |  |  |
| Minimum | 8.01 | 7.85 | 4.61 |
| Maximum | 13.20 | 11.60 | 13.65 |
| Median | 9.87 | 10.33 | 10.95 |
| Number of Imbalanced Schools | 7 (70\%) | 7 (70\%) | 0 (0\%) |
| Observed H value | 0.1640 | 0.0681 | 0.0121 |
| Simulated H values |  |  |  |
| Maintaining 15\% Rule |  |  |  |
| Mean of 10k replications | 0.0017 | 0.0013 | 0.0018 |
| 95\% Bounds | 0.0009-0.0028 | 0.0006-0.0021 | 0.0009-0.0029 |
| Maximum attainable | 0.1338 | 0.1760 | 0.1682 |
| Unconstrained Random Allocation |  |  |  |
| Mean | 0.0022 | 0.0018 | 0.0023 |
| 95\% Bounds | 0.0012-0.0036 | 0.0010-0.0028 | 0.0012-0.0036 |
| \% of replications yielding Imbalanced schools | 0 | 0 | 0 |

## Empirical Data and Methods

Data for this study are actual public school enrollments for the 551 public elementary schools of Riverside and San Bernardino Counties. From the National Center for Education Statistics electronic data
files reporting the number and ethnicity of school enrollments in all of the elementary schools within the two-county metropolitan statistical area under study, we accessed publicly available enrollments beginning with the year 1981. Earlier school ethnicity data are taken from four Office of Civil Rights reports published in 1968, 1969, 1971 and 1972. The school ethnicity data are not available prior to 1968, and were only episodically available prior to 1981 when annual data became available. Unfortunately school data prior to 1976 do not distinguish Hispanic students as a unique ethnic group. Beginning with 1976, we are able to divide the student populations into four ethnic groups: Caucasian ("White"), African American ("Black"), Latino/a ("Hispanic") and all others ("Asians \& Others"). All index values were calculated from these data using the equations specified above.

Since this study of segregation depends entirely on the comparability of school district jurisdictional boundaries over the study period, we note that boundary issues are virtually nonexistent. While the number of elementary schools has grown (and a significant number of schools have closed and disappeared over the four decade period of this study), the school district governance areas have been quite stable. In general, all elementary school district boundaries have been stable for the entire study period. While several unified school districts were formed over the years by breaking up large high school districts, the new districts were typically formed along the boundaries of the former elementary districts. This did not change the scope or location of governance agencies for the elementary schools which are the target of this study. A few rural elementary school districts merged with others in the formation of unified districts, typically because they were very small and hardly viable as independent districts (primarily in the Coachella Valley of southeastern Riverside County around the Salton Sea). Also, a few new unified districts separated from previously larger unified districts, typically because they were geographically remote and had experienced sufficient population growth to be independently viable (primarily in the Mojave desert region of northeastern San Bernardino County). All of these new districts are in rural areas that had never transported nor ever would be transporting students to other schools
due to long distances. In other words, changes in fragmentation at the elementary level have been marginal and the total number of jurisdictions in the two-county region has remained nearly constant at the current number of 53 unified or elementary districts covering all locations in Riverside and San Bernardino Counties.

## Measures Are Empirically Uncorrelated

As we reviewed earlier, different indices capture different ideas about the definition of segregation (also see, e.g., Johnston, Poulsen, \& Forrest, 2007; Massey \& Denton, 1988; Massey, White, \& Phua, 1996). To document similarities and differences among the indices, we examined their pairwise correlations. Since the axioms that define the various segregation indices measure ordinal, rather than interval relationships (e.g., see Frankel \& Volij, 2009), the mathematical functions used to calculate index values should be considered equivalent only if the values they assign to an array of school districts are perfectly rank-order correlated (i.e., have Spearman rather than Pearson correlations of 1.0). Comparing the exact numerical values across indices is not expected to produce perfect correlations, but comparing how the indices order the districts from more to less segregated tests their degree of similarity. Table 3 reports the Spearman rank-order correlations among all of the possible pairs of indices to determine if any two of them produce identical segregation orderings. With rank order correlations ranging from $+0.31\left(\mathrm{D}\right.$ with $\left.\mathrm{P}^{*}\right)$ to +0.97 ( $G$ with D ), it is clear that there are some very similar rank order values among the indices. None of these correlations are perfect, however, indicating that relying of different indices would yield different conclusions about the extent of segregation in various school districts. Theil's H, the Dissimilarity index, and the Gini coefficient are very close together, having 90\% to 95\% similar rankings (for similar analysis using Pearson product-moment correlations for two-group cases, see Massey \& Denton, 1988).

Table 2. Spearman rank-order correlations among the six indices (H, D, G, RIS, NWS, and P*).

|  | Segregation Index |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| Segregation Index | D | G | RIS | NWS | $\mathrm{P}^{*}$ |  |
| H |  |  | .952 | .958 | .751 |  |
| D |  | .401 | .314 |  |  |  |
| G |  | .975 | .672 | .424 | .312 |  |
| RIS |  |  | .708 | .453 | .329 |  |
| NWS |  |  |  |  | .262 |  |

Note: $\mathrm{N}=1,207$ district-year values.

Figure 2. Scatter plot matrix of each segregation index value against the other for the six calculated here (H, G, D, RIS, NWS, and $P^{*}$ ).


The relationships among the indices for all 1,207 available data points are presented as a matrix of scatter plots in Figure 2. Compared to all of the other pairs, only H with D and G, and D with G, display
relatively uninterrupted trends across the range of values. This is consistent with the H, D, and G pairs having the largest correlation coefficients in Table 2.

## Segregation Has Fluctuated Substantially and Appears Different with Different Measures

When we examine longitudinal data for the Riverside/San Bernardino metropolitan area the divergence in definitions and resulting calculations for these six desegregation measures is easily recognized. Figure 3 graphically presents the yearly central tendency values for all six segregation indices for each of the years for which data are available. To overlay the index values and make their differences easier to see, the median values for Theil's H are scaled on the left axis, the much larger values for the other indices are scaled by the figure's right side axis. The median values are plotted for each index, except the racially isolated (non-white) schools (NWS) proportion. In each year there are generally only a few ethnically isolated schools reaching the $90 \%$ or more non-white, hence the NWS measures are highly skewed (with a median value of 0 for all years). Hence, the mean rather than the median is a better way to track changes in this index.

The most obvious message in Figure 3 is that the median level of segregation dropped substantially following 1968 for all measures. After 1971, however, the question of whether segregation became more or less severe clearly depends on how segregation is defined and measured. Measured by the three highly correlated indices ( $H, D$ and $G$ ), the degree of segregation either continued to drop or stabilized at values appreciably less than the 1968 value until about 1990. Measured by the three isolation and concentration indices ( $\mathrm{P}^{*}$, RIS and NWS), however, segregation began to get worse starting in 1972 and has become progressively more sever throughout the last 35 years. The $P^{*}$ median steadily increased each year beginning in 1972. The NWS mean remained very close to zero until about 1990, but has been rising at an increasing rate since then. The median number of imbalanced schools (the RIS median) has had some fairly sharp year to year fluctuations, but there has been a clear upward trend in this index throughout the period from 1972 to 2007. Measured by the median values of $H, D$ and $G$,
essentially all of the desegregation gains made in the 1980s and early 1990s has been lost - though the districts are substantially more integrated than they were in the 1960s and 70s. The RIS and NWS values in 2007 were significantly higher than in any prior year, with the trend looking decidedly like even worse days are ahead. The three evenness indicators ( $\mathrm{H}, \mathrm{D}$ and G ) are also showing an upward trend, though not with the steepness of $P^{*}$ and NWS. The six indices show no agreement about when the general level of segregation was at an all-time low. The RIS, NWS, and $\mathrm{P}^{*}$ indices would put the all-time lows during the 1970s, whereas H, D, and G put the all-time lows right around 1990.

Figure 3. Trends in the central tendency of school district segregation index values (H, D, G, P*, RIS, and NWS) among Riverside County and San Bernardino County districts, 1968-2007.


We also note that, with the exception of the years 1969 and 1971, the RIS median has always been greater than 0.10 . This means that more than $10 \%$ of the elementary schools in at least half of the districts in this study were ethnically imbalanced under the $15 \%$ rule. By 2007, this number had more than doubled - at least $25 \%$ of the schools in more than half the districts in this metro area were ethnically imbalanced in that year. If we ignore district boundaries and just assess the percentage of the

551 schools in the region we find that about $30 \%$ of all schools in the region have been ethnically imbalanced every year, throughout the four-decade study period.

## Regional Segregation Has Moved from Within to Between Districts

Some of the trend differences observed in Figure 3 can be explained by an overall shift in the ethnic composition of districts within the two-county region. Partitioning Theil's H into within and between district components reveals that the desegregation era concentrated in the 1960's and 70's ushered in a dramatic decline in between-school segregation and had only modest impact on between-

Figure 4. Theil's H measuring between elementary schools and between districts segregation for public school districts in the Riverside-San Bernardino metropolitan area, 1968-2007.

district segregation (tracked by the red line on Figure 4). Measured by Theil's H, as within school district segregation was declining sharply, between district segregation grew significantly, and then declined somewhat in the 1980's only to begin a steady climb between 1990 and 2005. In sum, as districts were
substantially desegregating their schools, regional population shifts were exacerbating between district segregation. And since 1990 both within and between district segregation has been steadily rising. Segregation Is More Intense between White and Non-White Students than Among Non-Whites

Non-white population groups have grown so much more rapidly in some districts that racially isolated (non-white) schools have emerged where they had never been seen before. The ethnic isolation NWS index values are rising primarily due to Hispanic in-migration into some school districts and substantial declines in the white residential populations in the region's largest cities. The partitioning of Theil's H into between and among racial/ethnic groups in this region was investigated by Mitchell, Mitchell, and Batie (2009, see Figures 16 through 18) and that analysis will not be repeated here. They found that the largest Theil's H values were almost always between white and non-white students with substantially smaller proportions of total segregation attributable to segregation among the non-white groups. The white/non-white proportion of segregation was highest at about $70 \%$ in the late 1960 s and has been mostly in the middle $60 \%$ range ever since. The black share was also highest in the late 1960s, but it has steadily declined to a third its earlier level by the late 2000s. The Hispanic share, on the other hand, was at its lowest in the 1960s and has increased steadily to its highest levels in the late 2000s (reaching the same level as that for whites). In short, white and Hispanic students have become more separated both within and between school districts—more between than within districts.

## Conclusion

This paper has tackled three important aspects of the measurement of elementary school desegregation. First, we reviewed the character and meaning of alternative measures of school segregation and noted that they are based on different definitions of how to monitor the transition from complete segregation (when all students attend schools with a single ethnic group) to complete desegregation (when all schools have a student mix that exactly matches the mix of ethnic groups in the entire school district). We noted that the indices agree in assigning a zero when all schools are fully
integrated and a value of one when all schools are completely segregated. We found, however, that the major indices (H, D, G, $\mathrm{P}^{*}$, RIS and NWS) follow different trajectories, meaning that no two indices are perfectly correlated in real-world assessments of school desegregation as schools move between the extremes of apartheid and integration.

Second, using Monte Carlo random assignment of students to schools, we found that the expected values for Theil's $H$, the most mathematically defensible measure of segregation levels, are very much smaller than the values for this index found in a very large sample of school composition measures in the nation's $13^{\text {th }}$ largest metropolitan region. Thus, we concluded that school enrollment patterns over a 40 year period and a large metropolitan region are very far from randomly distributed. Third, we noted that, while intra-district segregation declined sharply under pressure from the civil rights movement in the 1960s and 70s, it has not improved much since about 1990. Indeed, withindistrict school-level segregation has been creeping up again since that time. Additionally, in this metropolitan area, the between district segregation levels have been rising steadily since 1990. These two factors taken together mean that the number of racially isolated, non-white schools has been rising at an increasing rate throughout the 1990s and since 2000. In 2007, when our data collection ended, fully $20 \%$ of the schools in this metropolitan region had become racially isolated through the combined effects of inter-district and intra-district ethnic segregation.

## Perfect Desegregation Is Not a Realistic Goal

We note in conclusion, that perfect desegregation is not a realistic goal. This fact has been recognized throughout the history of judicial and political efforts to create racial and ethnic balance in the nation's schools. Initially, the goal was only to insure that no school would be deemed racially isolated by being composed of $90 \%$ or more of non-white students. As the political system struggled for a more complete desegregation, states like California settled on an ethnic imbalance criterion-insisting that no school should have an enrollment that deviated more than $15 \%$ from district-wide ethnic
proportions. California settled on $15 \%$, however, some localities were more stringent and some more lenient. All agreed, however, that the political and economic costs of perfect desegregation were prohibitively high, and probably not essential to meet the social and cultural purposes of integration. What has been largely unrecognized in debates over how to measure school desegregation is the fact that under conditions of partial desegregation, the various measures of segregation reach importantly different conclusions about how close to the criterion of substantially desegregated any given school or district might be.

## Information Based Measures Are Most Defensible, but ...

The picture is even more complex when we consider the virtues and problems of various segregation measures as they track intermediate levels of desegregation. Theil's information index (H) has been shown to be mathematically superior. It can be partitioned to separate the degree of segregation attributable to differences in sub-regions within a larger region (like the degree of segregation between districts in a metropolitan area). As we stated previously, H is unique in helping to identify the victims of segregation and allocate political and administrative responsibility for providing remedies.

Theil's H can also be partitioned to distinguish the extent to which measureable segregation is being created by the isolation of one or another specific ethnic group, rather than being produced by all groups equally. This mathematical superiority makes H a very attractive measurement tool, but it provides very little guidance to administrators or policy makers because it does not clearly specify how many or which students need to be reassigned to reduce the level of segregation. Moreover, this index follows a curvilinear path away from perfect desegregation and thus makes small improvements among highly segregated schools seem large compared to larger improvements made by more fully desegregated systems. Most dissatisfying, however, is that there is no numerical value of H that can be established as a threshold below which reasonable levels of desegregation will have been achieved. As
discovered here, a district may have multiple ethnically imbalanced schools while its H value is, nonetheless, much smaller than the maximum attainable in satisfaction of the $15 \%$ rule.

## Dissimilarity Is the Easiest to Understand and Apply, but ...

Contrasting with Theil's H, the Dissimilarity index has two important virtues. It reports just what proportion of a school district's population is involved in the measured segregation, and it follows a linear path between full desegregation and full segregation thus giving movement at any point along the continuum a similar statistical meaning. These virtues do not, however, make up for the lesser mathematical elegance and lack of partitionability of this index. Hence, it is probably important for studies of school segregation to routinely report both of these measures.

Isolation Can Be Hidden in Broader Averaging Measures

A review of the ethnic isolation type indices makes it clear that neither H nor D will reliably identify school districts in which one or more schools have been allowed to become ethnically imbalanced or racially isolated. Indeed, in a very large sample of school district assessments we found that measured values of H and D that are substantially smaller than the maximum values that could be achieved by districts meeting the $15 \%$ criterion for acceptable desegregation levels nevertheless contain multiple schools that have been allowed to become ethnically imbalanced. Indeed, we even find some school districts with H values smaller than the maximum that can be attained without violating the $15 \%$ criterion nevertheless contain one or more schools that qualify as ethnically isolated (using the $90 \%$ non-white criterion). Thus, for monitoring of school district progress toward desegregation it is important to monitor H to be sure measures are reliable and responsibility portioned, to monitor D to learn how much needs to be done, and to monitor RIS and NWS to find out if specific schools are seriously imbalanced.

## There Are Issues That Need Further Development

In order to develop a complete picture of ethnic segregation in the nation's public schools there are at least two issues that need further theoretical and methodological work. The first is assessing geographic concentrations that may be making desegregation prohibitively expensive, and the second is development of measures that allow us to compare the levels of segregation between school systems with different numbers of schools and different sizes of schools.

## The Assessment of Geographic Concentration Needs a Good Measure

One measure that is not easily produced with current methods, but for which the needed data are available in raw data form is an assessment of how far students might have to be moved to produce adequate levels of desegregation. While the identification of schools that are ethnically imbalanced has become relatively straightforward, there are no easy measures of the extent to which nearby schools are also ethnically imbalanced, making student movement time consuming and costly. Particularly in the nation's metropolitan centers, ethnic ghettos tend to be much larger than the size of an elementary school catchment area and hence impose serious constraints on the ability to produce politically acceptable levels of desegregation. In some places ethnic concentrations in charter or magnet schools are being criticized without any realistic assessment of what would be required to generate balanced enrollments.

## Changes in District Size and Composition Need Clearer Interpretation

Finally, we note that attempts to chart a school district's progress toward desegregation using currently available measures is rendered somewhat unreliable by the fact that districts open new schools, close old ones, redraw catchment area boundaries and undergo population shifts that, in combination change the number of schools and substantially alter the size of their schools. Current measures are not immune to misleading assessments when trying to compare data sets representing different mixes of school numbers and sizes, which are occurring simultaneously with changes in the
overall ethnic composition of the district. This issue needs further measurement attention from research scholars with sophisticated mathematical skills and appropriate sensitivities to how the development of new indices might be altering the conceptual definition of school desegregation-especially in the middle ranges of desegregation which appear to be the best that can possibly be achieved within the levels of political commitment and financial support that schools can expect for the foreseeable future.

## Notes

1. Only the multigroup Theil's H obeys the transfer axiom (Reardon \& Firebaugh, 2002), which means that except for H no multigroup segregation index unerringly makes a comparison when school enrollment levels change from one measurement occasion to another. At least in principle, failure to obey the transfer axiom is a devastating shortcoming for any index that is used to monitor the level of segregation over time, though the magnitude of the error for this (mis)application is unknown at this time.

## References

Berger, M. A. (1984). Neighborhood schools: The new(?) legal response to enrollment decline and desegregation. Urban Education, 18(1), 7-28.

Bischoff, K. (2008). School district fragmentation and racial residential segregation: How do boundaries matter? Urban Affairs Review, 44(2), 182-217.

Clotfelter, C. T. (1979). Urban school desegregation and declines in white enrollment: A reexamination. Journal of Urban Economics, 6(3), 352-370.

Clotfelter, C. T. (1999). Public school segregation in metropolitan areas. Land Economics, 75(4), 487-504.
Clotfelter, C. T. (2001) Are whites still fleeing? Racial patterns and enrollment shifts in urban public schools. Journal of Policy Analysis and Management, 20(2), 199-221.

Clotfelter, C. T. (2004) After Brown: The rise and retreat of school desegregation. Princeton, NJ: Princeton University Press.

Clotfelter, C. T., Vigdor, J. L., \& Ladd, H. F. (2006). Federal oversight, local control, and the specter of "resegregation" in southern schools. American Law and Economics Review, 8(2), 347-389.

Conger, D. (2005) Within-school segregation in an urban school district. Educational Evaluation and Policy Analysis, 27(3), 225-244.

Duncan, O. D., \& Duncan, B. (1955). A methodological analysis of segregation indexes. American Sociological Review, 20(2), 210-217.

Dye, T. R. (1968). Urban school segregation: A comparative analysis. Urban Affairs Quarterly, 4(2), 144165.

Dziuban, C. D., \& Esler, W. K. (1983). Some relationships among indices of school desegregation. American Education Research Journal, 20(1), 113-122.

Edelman, M. W. (1973). Southern school desegregation, 1954-1973: A judicial-political overview. The Annals of the American Academy of Political and Social Science, 407, 32-42.

Frankel, D. M., \& Volij, O. (2009, March 5). Measuring school segregation. Ames, IA: Department of Economics, Iowa State University. Retrieved March 15, 2010, from http://econ2.econ.iastate.edu/faculty/frankel/segRvsnJET.pdf

Frankenberg, E., Lee, C., \& Orfield, G. (2003, January). A multiracial society with segregated schools: Are we losing the dream? Cambridge, MA: The Civil Rights Project, Harvard University.

Giles, M. W. (1977). Racial stability and urban school desegregation. Urban Affairs Quarterly, 12(4), 499510.

Goodman, L. A., \& Kruskal, W. H. (1959). Measures of association for cross classifications. II: Further discussion and references. Journal of the American Statistical Association, 54(285), 123-163.

Logan, J. R., \& Oakley, D. (2004, January 28). The continuing legacy of the Brown decision: Court action and school segregation, 1960-2000. Albany, NY: Lewis Mumford Center for Comparative Urban and Regional Research, University at Albany. Retrieved April 6, 2010, from http://www.s4. brown.edu/schoolsegregation/reports/brown/Brown_report_1_28.pdf.

Logan, J. R., Oakley, D., \& Stowell, J. (2006). Resegregation in U.S. public schools or white decline? A closer look at trends in the 1990s. Children, Youth and Environments, 16(1), 49-68. Retrieved April 6, 2010, from http://www.colorado.edu/journals/cye.

Logan, J. R., Oakley, D., \& Stowell, J. (2008). School segregation in metropolitan regions, 1970-2000: The impacts of policy choices on public education. American Journal of Sociology, 113(6), 1611-1644.

Johnston, R., Poulsen, M., \& Forrest, J. (2007). Ethnic and racial segregation in U.S. metropolitan areas, 1980-2000: The dimensions of segregation revisited. Urban Affairs Review, 42(4), 479-504.

Massey, D. S., \& Denton, N. A. (1988). The dimensions of racial segregation. Social Forces, 67(2), 281315.

Massey, D. S., White, M. J., \& Phua, V.-C. (1996). The dimensions of segregation revisited. Sociological Methods \& Research, 25(2), 172-206.
Mele, A. (2009, February 7). Poisson indices of segregation. (MPRA Paper No. 15155). Munich, Germany: Munich Personal RePEc Archive (MPRA). Retrieved March 15, 2010, from http://mpra.ub.unimuenchen.de/15155/1/MPRA_paper_15155.pdf

Mitchell, D. E., Batie, M., \& Mitchell, R. E. (2010). The contributions of school desegregation to housing integration: Case studies in two large urban areas. Urban Education, 45(2) 166-193.

Mitchell, D. E., Mitchell, R. E., \& Batie, M. (2009, April 15). Multiethnic metropolitan school desegregation and housing integration. Paper presented at the annual meeting of the American Educational Research Association, San Diego, California.

Orfield, G. (1983). Public school desegregation in the United States, 1968-1980. Washington, DC: Joint Center for Political Studies.

Orfield, G., \& Lee, C. (2007, August). Historic reversals, accelerating resegregation, and the need for new integration strategies. Los Angeles: The Civil Rights Project/Proyecto Derechos Civiles, University of California, Los Angeles.

Orfield, G., \& Yun, J. T. (1999, June). Resegregation in American schools. Cambridge, MA: The Civil Rights Project, Harvard University.

Reardon, S. F. (2006). A conceptual framework for measuring segregation and its association with population outcomes. In J. M. Oakes \& J. S. Kaufman (Eds.), Methods in social epidemiology (pp. 169-192). San Francisco: Jossey-Bass.

Reardon, S. F., \& Firebaugh, G. (2002). Measures of multi-group segregation. Sociological Methodology, 32, 33-67.

Reardon, S. F., \& Yun, J. T. (2001). Suburban racial change and suburban school segregation, 1987-1995. Sociology of Education, 74(2), 79-101.

Reardon, S. F., \& Yun, J. T. (2005). Integrating neighborhoods, segregating schools: The retreat from school desegregation in the South, 1990-2000. In J. C. Boger \& G. Orfield (Eds.), School resegregation: Must the South turn back? (pp. 51-69). Chapel Hill, NC: University of North Carolina Press.

Reardon, S. F., Yun, J. T., \& Eitle, T. M. (2000). The changing structure of school segregation: Measurement and evidence of multi-racial metropolitan area school segregation, 1989-1995. Demography, 37(3), 351-64.

Reardon, S. F., Yun, J. T., \& Kurlaender, M. (2006). Implications of income-based school assignment policies for racial school segregation. Educational Evaluation and Policy Analysis, 28(1), 49-75.

Rossell, C. H. (1985). Estimating the net benefit of school desegregation reassignments. Educational Evaluation and Policy Analysis, 7(3), 217-227.

Rossell, C. H., \& Armor, D. J. (1996). The effectiveness of school desegregation plans, 1968-1991. American Politics Quarterly, 24(3), 267-302.

SAS Institute Inc. (2008). Base SAS ${ }^{\circledR} 9.2$ procedures guide: Statistical procedures. Cary, NC: SAS Institute Inc.

Smith, S. S., \& Mickelson, R. A. (2000). All that glitters is not gold: School reform in CharlotteMecklenburg. Educational Evaluation and Policy Analysis, 22(2), 101-127.

Stearns, L. B., \& Logan, J. R. (1986). Measuring trends in segregation: Three dimensions, three measures. Urban Affairs Quarterly, 22(1), 124-150.

Valencia, R. R., Menchaca, M., \& Donato, R. (2002). Segregation, desegregation, and integration of Chicano students: Old and new realities. In R. R. Valencia (Ed.), Chicano school failure and success: Past, present, and future (2nd ed., pp. 70-113). London: RoutledgeFalmer.

Welner, K. G. (2006). K-12 race-conscious student assignment policies: Law, social science, and diversity. Review of Educational Research, 76(3), 349-382.

Wollenberg, C. (1974). Mendez v. Westminster: Race, nationality and segregation in California schools. California Historical Quarterly, 53(4), 317-332.

Zoloth, B. S. (1976). Alternative measures of school segregation. Land Economics, 52(3), 278-298.

Legal Cases
Brown v. Board of Education of Topeka, 347 U.S. 483 (1954).
Brown v. Board of Education of Topeka, 349 U.S. 294 (1955).
Green v. County Sch. Bd. of New Kent County, 391 U.S. 430 (1968)
Jackson v. Pasadena City School Dist., 59 Cal.2d 876 (1963)
Keyes v. School Dist. No. 1, 413 U.S. 189 (1973)
Mendez v. Westminster, 64 F. Supp. 544 (1946)
Milliken v. Bradley, 418 U.S. 717 (1974)
Milliken v. Bradley, 433 U.S. 267 (1977)
National Assn. for Advancement of Colored People v. San Bernardino City Unified Sch. Dist., 17 Cal.3d 311 (1976)

Santa Barbara Sch. Dist. v. Superior Court, 13 Cal.3d 315 (1975)
Swann v. Charlotte-Mecklenburg Bd. of Educ., 402 U.S. 1 (1971)

