
PUB DATE
2 Dec 75

NOTE

EDRS PRICE
MF-$0.83 HC-$1.67 plus Postage.

EDUCATIONAL POLICY; EDUCATIONAL RESEARCH; ERROR PATTERNS; EVALUATION METHODS; FORMAL CRITICISM; INTEGRATION EFFECTS; INTEGRATION METHODS; MINORITY GROUP CHILDREN; POLICY FORMATION; RESEARCH METHODOLOGY; RESEARCH PROBLEMS; SCHOOL INTEGRATION; SOCIAL CHANGE; STATISTICAL ANALYSIS; STATISTICAL BIAS; STATISTICAL DATA; THEORETICAL CRITICISM

ABSTRACT
Suggesting that James Coleman's paper on massive school desegregation reveals methodological flaws of such magnitude that they raise serious questions as to the validity of the conclusions, this paper addresses a full sequence of perceived methodological errors found in the Coleman document, but does not dismiss the conclusions based upon initial errors, no matter how cogent they be. This approach makes the implicit assumption that, at each step, all previous operations upon the data are both well reasoned and technically correct. As the critique progresses, it becomes obvious to what extent this assumption is incorrect. The first point made concerns the data base used, and, by implication, the range of variables considered for inclusion in the various regression models used. The second point relates to the unit of analysis employed. Another potential problem concerns the partitioning of the data into two categories based upon school district size. The treatment (or lack of treatment) of the standard errors of the regression coefficients is also criticized. Finally, one of the more misleading of the methodological errors is reflected in the attempt to control for the unique characteristics of the cities through the use of dummy variables. (Author/AM)
In numerous press conferences, interviews, Congressional hearings and, most recently, in the pages of the Phi Delta Kappan Dr. James Coleman has made the rather forceful statement that his research shows massive school desegregation efforts to be self-defeating. It is his position that such attempts simply hasten the departure of white children from the public schools of the involved districts. While such statements have struck a responsive chord in a large section of Americans, contrary to his claim they are not very well grounded in empirical facts. The analyses that have, until recently, been unavailable to the general public and the research community are now set forth in "Trends in School Segregation, 1968-73". A careful perusal of this document reveals to the reader methodological flaws of sufficient magnitude to raise serious questions as to the validity of the conclusions.

What now follows represents this author’s attempt to offer a critique of the above mentioned paper. As such, no claims are made to its completeness; there could well be errors of omission or commission which would serve to further weaken the inferential chain welded by Coleman et al. What is done by the author attempts to address a full sequence of perceived methodological errors rather than dismissing the conclusions based upon initial errors no matter how cogent. This approach, therefore, makes the implicit assumption that, at each step, all previous operations upon the data have been both well reasoned and technically correct. As the critique progresses it will become painfully obvious to what extent this assumption is, in fact, incorrect.

Presented before a Seminar on Dr. Coleman’s research on Desegregation and White Flight held at Michigan State University, December 2, 1975.
The first point which needs be made concerns the data base used by the authors and, by implication, the range of variables considered for inclusion in their various regression models. Their sole source of information appears to have been data supplied by school districts to the Department of Health, Education and Welfare. The variables on which information was supplied relate solely to public school attendance: total number of students, total white, total black, their respective proportions in the various districts and schools, and the changes taking place in them from year to year. While, in and of themselves, they present an interesting view of changes taking place in public education, they cannot be relied upon to furnish us with the degree of understanding necessary for us to address the causes of decreasing white enrollment.

Even if we are to make the assumption, as do Coleman et al, that this phenomenon reflects white residential relocation ("white flight"), we must ask ourselves what variables would likely be related to such an activity. Likely sets of variables would, of necessity, include factors related to economics, demographics, crime and violence, and the quality of the environment to name but a few. While public school desegregation deserves its place as an educational factor potentially affecting white school attendance, focusing upon such a phenomenon as the primary predictive variable would appear to be both shortsighted and highly likely to lead to incorrect inferences about the true nature of the problem.

The second methodological point relates to the unit of analysis employed by Coleman et al. Based on what is reported as the number of observations and degrees of freedom one must assume that the basic unit of analysis is each year's data for each city, (or the between year's differences to be more accurate). Apparently the roughly 5 between-adjacent-pairs-of-years differences for each city are merely thrown together for use in the regression analysis with the specific city identification completely disregarded.
This would seem to be completely inappropriate. A more reasonable procedure would have been to have used each city as the unit of analysis, forming some measure of relationship between change in white attendance and change in segregation within each city based on the time points for that city alone. These could then have been averaged across all cities with appropriate confidence intervals established based on correct degrees of freedom.

Yet another questionable approach is the pooling of data from the various years to predict the proportion change in white attendance. This would seem to be inappropriate in that individual years are disregarded with data arising from one pair of years being considered equally with data arising from any other pair of years. Since it would seem that the author's intentions were to investigate trends in school segregation, such an approach ignores a key variable in their study.

Another potential problem has to do with the partitioning of the data into two categories based upon school district size. Coleman et al justify this partitioning "because of the indications that response to desegregation differs considerably in the very large cities from the response in the smaller ones". It would have been quite easy for such an assumption to have been statistically tested by pooling all data, doing the regression analysis, then introducing a dummy variable to see if a significant gain in the $R^2$ was produced. If not, the coefficients could then have been estimated with smaller standard error due to the larger number of degrees of freedom for error.

An even more disturbing practice has to do with their treatment (or lack of treatment) of the standard errors of their regression coefficients. In most instances the authors failed to temper substantive interpretation of regression coefficients in the light of their standard errors. Where standard errors are sufficiently large as to make tenable the null hypothesis that a
particular regression coefficient is not different from zero, the authors should have eliminated the variable from the equation, rerun the analysis, and based subsequent use of the equation on the more parsimonious model. In Table 14 and the use of the results displayed, there is a glaring example of this problem since many of the coefficients are not substantially larger than their standard errors, (or are, in fact, smaller). In producing their estimates for various changes in their dependent measures they again failed to take into account the standard errors of their coefficients. Had confidence intervals been generated about their estimated loss of white students, it seems inevitable that 0% loss would be included.

A fairly minor error of omission takes place in the author's caution with respect to the interpretation of the regression coefficients presented in Table 14. The authors state that there is some difficulty in interpreting the actual coefficients in the regression equations citing problems associated with multi-collinearity. While this may be true, the authors should present empirical evidence in the form of a correlation matrix of the variables to make such an interpretation more plausible as well as to give the reader more information about what is actually taking place in the data.

In arriving at estimates for regression coefficients discussed later on, the authors state that they averaged the first three coefficients common to four time-lagged regression models presented earlier. Such an averaging procedure is completely unjustifiable. For each model the first three estimates are conditioned upon the inclusion of the additional terms, and unless the first three terms are statistically independent of all subsequent terms, they do not estimate the same quantity. Their statement that such an analysis cannot provide a conclusive answer is grossly understated in the light of their erroneous methods.
One of the more misleading of the methodological errors is reflected in their attempt to control for the unique characteristics of the cities through the use of dummy variables. Even if the criticisms mentioned with respect to the unit of analysis are ignored (and they must be to allow for this particular analysis), a very stringent assumption must be met if the author's use of dummy variables for cities is to do as much for this analysis as they seem to imply. Since the values of the dummy variables are constant across all six years for each individual city, we must assume that the values for all other possibly relevant variables remain constant. This assumption seems incredibly farfetched on its face. In addition, the authors present no empirical evidence which would justify it. Clearly, these cities have not only their own particular rate of economic activity, crime, etc., but their own particular year-to-year changes in these rates—change which could be related as strongly to changes in white school attendance as are changes in the segregation index.

Finally, in discussing predicted white school attendance across time, the authors chose to focus upon predictions arising from Equation 3 rather than those from Equation 1 (See Table 14). Even here, the authors fail to provide confidence intervals on their estimates, a procedure which would seem especially important since four of the eight variables in equation 3 have coefficients less than, or only slightly larger than, their standard errors. In the light of the instabilities likely to result from such a situation, it would seem more appropriate to focus upon the projections based on equation 1—projections which show a difference of only 3% with desegregation compared to no desegregation.

Thus it may be seen that the work reported by Coleman et al suffers from a sequence of serious methodological flaws. Their failure to replicate the findings of other researchers more equivocal with respect to the impact of desegregation upon white enrollment may be due, in large part, to the afore mentioned problems. While no firm statement can be made to the accuracy of their conclusions, one thing which can be stated is that they have wandered far from their original data in coming to those conclusions.
Reference

Coleman, James S., Sara D. Kelly and John A. Moore

### Table 14.
**Regression Coefficients for Analyses of White Student Loss to Central Cities**

**Largest 21** | **Next 46**
--- | ---
\( \Delta R \) & .277(.062) & .091(.031) \\
Prop. black & -1.35(.028) & -0.86(.151) \\
\( \ln N \) & .001(.001) & -0.047(.011) \\
Constant & .503 & .503 \\
\( R^2 \) & .28 & .24 \\
Number of observations & (103) & (239) \\
Including inter-district segregation in SMSA, and interaction of desegregation with South

**Equation 2**

\( \Delta R \) & .195(.158) & .008(.151) \\
Prop. black & -0.047(.041) & -0.033(.019) \\
\( \ln N \) & .007(.008) & -0.042(.011) \\
R SMSA & -1.62(.052) & -1.09(.024) \\
\( \Delta R x S \) & .144(.172) & .122(.152) \\
Constant & .450 & .450 \\
\( R^2 \) & .35 & .29 \\
Including interactions of desegregation with proportion black and inter-district segregation, and also including South as dummy variable

**Equation 3**

\( \Delta R \) & -0.460(.187) & -0.147(.173) \\
Proportion black & .050(.037) & -0.027(.022) \\
\( \ln N \) & .003(.007) & -0.041(.011) \\
R SMSA & -0.208(.045) & -0.101(.029) \\
\( \Delta R x S \) & .146(.201) & .108(.165) \\
\( \Delta R x p b l \) & 1.774(.313) & .406(.254) \\
\( \Delta R x S M S A \) & .544(.501) & .664(.385) \\
South & -.006(.010) & -.001(.007) \\
Constant & .43337 & .43337 \\
\( R^2 \) & .60 & .32 \\

Reproduced from "Trends in School Segregation, 1968-73" - page 59