DOCUMENT RESUME

ED 135 850 TM 006 080

AUTHOR Gastright, Joseph F.
TITLE Some Empirical Evidence on the Comparability of School Unit Residuals Based on Achievement and Non-Achievement Variables.
PUB DATE [Apr 77]
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement Tests; Elementary Education; *Elementary Schools; *Measurement Techniques; *Multiple Regression Analysis; Predictor Variables; *Program Effectiveness; *School Demography; *Schools; Student Mobility

ABSTRACT

School unit residuals from multiple regression predictions of sixth-grade reading scores were developed for a population of 66 elementary schools, using Dyer's Method II (matched longitudinal data) and Method III (unmatched longitudinal data). The input variable sets were (1) previous achievement data alone, (2) achievement data and nonachievement data, and (3) nonachievement data alone. Residuals for the various methods and sets of input variables were correlated to study their comparability. Residuals from Method II and III correlated highly. Residuals based on nonachievement variables were unrelated to those based on achievement alone. (Author)

SOME EMPIRICAL EVIDENCE ON THE COMPARABILITY OF SCHOOL UNIT RESIDUALS BASED ON ACHIEVEMENT AND NON-ACHIEVEMENT VARIABLES

by

Joseph F. Gastright
Program Evaluation Branch
Cincinnati Public Schools

Presented at annual meeting of the
American Educational Research Association
New York, New York

April, 1977

Some Empirical Evidence on the Comparability of School Unit Residuals Based on Achievement and Non-Achievement Variables

JOSEPH F. GASTRIGHT
Cincinnati Public Schools

The evolving interest in the objective assessment of school performance has raised an embarrassing issue for educators. The truth is that it is not that obvious how school performance should be quantified in order to establish output accountability. Methods which were acceptable in the past as descriptions of school status have proved unacceptable as measures of school performance.

Until recently the question of school quality was handled by accreditation. This method focused on the resources and curricula of the school as though these factors were ends in themselves. Schools were periodically certified by a site visit team which concentrated on such items as the number of books in the library, the variety of course work, and the academic background of the teaching staff.

Achievement test scores, if they were used at all, were used to characterize the population served by the school. Critics of school performance have dismissed this process as subjective and inherently predisposed to over rate the accomplishment of schools.

These same critics have seized upon the standardized survey tests of academic achievement, given in some form in virtually all schools, as the appropriate objective measure of school effectiveness. This has often led to the unreasonable conclusion that the discrepancy between a school's achievement test scores and those of another school, or those of the national norm group, is an unbiased measure of the effectiveness of the school.

Quite clearly any achievement comparison which ignores both the ability levels of students entering the school and the surrounding conditions under which the school operates must be a fallacious measure of acceptable performance.

Dyer (1966) has suggested that a more rational estimate of school effectiveness might be based on the discrepancy between the observed output achievement scores of a school and multiple-regression predictions of its output achievement test scores predicted from measures of previous achievement and measures of hard-to-change conditions which affect learning. Dyer, Linn and Patton (1969) provided empirical evidence on the stability and comparability of residuals based on previous achievement alone obtained using four different methods of aggregating the achievement data.

Method I utilized the regression of individual student output achievement on student input achievement using a student sample identical at two grade levels (matched-longitudinal sample). Method II utilized the regression of mean school system output on mean school system input for the same matched-longitudinal sample of students. Method III utilized the regression of mean school system input for all students available at those points in time (unmatched-longitudinal sample). Method IV utilized regression of mean school system output on the concurrent school system mean of the earlier grade level (cross-sectional sample).

Dyer, et. al. (1969), using the fifth grade achievement scores to predict eighth grade achievement scores (Iowa Test of Basic Skills), concluded that Methods I and II were essentially interchangeable, but not comparable to Methods III and IV. Operating under the assumption that the methods utilizing matched student samples were intrinsically superior to the

others, they concluded that Methods III and IV did not produce residuals which were sufficiently comparable to those from Methods I and II to serve as reasonable substitutes for them.

Forsythe (1973) provided some evidence on a different kind of stability, the consistency of residuals for consecutive classes in the same school. He randomly sampled 50 students from each of 320 schools in Iowa and utilizing Dyer's Method II, predicted mean school twelfth-grade achievement scores by using mean school ninth-grade achievement scores (Iowa Test of Basic Skills). The multiple correlation coefficients reported by Forsythe are very consistent with those reported by Dyer. However, the inter-correlations between residuals for the consecutive years (median r=.28) were considerably lower than the random halves correlations reported by Dyer.

Gastright (1974) reported that residuals from regression analyses based on school unit data which included school background variables (Dyer's Method III) were more stable from year-to-year than those reported by Forsythe using the matched student samples. With the exception of Gastright's study, none of the published literature on longitudinal models studied the use of non-achievement variables on the comparability of residuals. If, as had been proposed by Coleman (1966) and Jencks (1972), these variables are highly related to achievement, then their absence in multiple regression studies of school performance could invalidate the results.

Convey (1975) used a simulated data base including both previous achievement and SES variables to test the validity of three methods of obtaining residuals. He concluded that residuals based on school unit data were superior in a cost-effectiveness sense. Convey also concluded that non-achievement variables should be assessed on the basis of theory, previous

research, or insight rather than on an empirical basis in order to determine the relative effectiveness of schools.

With the exception of Gastright (1974) and Convey (1975) there are few reports which include non-achievement variables in the determination of school effectiveness. Therefore, a major component in Dyer's original proposal for determining school effectiveness has received little attention in the literature. The purpose of this study is to examine the comparability of school unit residuals obtained using Dyer's Method II and Method III when these residuals are based on: a) previous achievement alone, b) previous achievement and school background variables, and c) school background variables alone.

Procedures

The data on school units were available on sixty-seven elementary schools in the Cincinnati Public School system. The children in these schools were tested in the third grade in the spring of 1970 with the Metropolitan Achievement Test, Primary II level, and tested again in the spring of 1973 with the Metropolitan Achievement Test, Intermediate level.

Data on over 800 non-achievement variables were available for each of the elementary schools. This data is compiled from the census reports, city agencies, and various departments within the school system.

The individual student results at both the third and sixth grades were matched on student name within each school to obtain the matched longitudinal population for Dyer's Method II. After this matching, the remaining cards were alphabetized for the total system and matched to obtain the population of students who changed schools within the system between 1970 and 1973. This second matched longitudinal population was

sorted by school, based on their location in the sixth grade, and added to the first set. The total matched student population for this study for a particular school includes both those students who remained in the school and those students who entered the school from within the system between 1970 and 1973.

The remaining set of unmatched third grade cards, by school, was made up largely of students who moved out of the system between 1970 and 1973. However, small percentages of these students may have either failed a grade between the third and sixth grades, or missed the sixth grade testing for some other reason.

The remaining set of unmatched sixth grade cards, by school, was made up largely of students who moved into the system between 1970 and 1973. However, small percentages of these students may have missed the third grade testing for some reason.

The test scores at both the third and sixth grades were reported as raw scores. They were converted to the equivalent standard scores on the Metropolitan Achievement Test Battery for all analyses in this study. The means of both the matched longitudinal sample for each school and the unmatched longitudinal sample for each school were computed and used in all subsequent analyses.

Thirty-seven input variables were selected from the eight hundred specific variables available on each school. This selection was on the basis of previous research, continuing availability, and the estimated accuracy of the data.

The variables in the School Information System have been analyzed previously to determine those variables which are most valuable as correlates of achievement. These studies have identified a subset of interpretable and accurately collected variables which represent all of the major categories reported in the system. From this much shorter list, those variables which are only sporadically available were eliminated. The final list of potential input variables contained thirty-seven achievement, mobility, and demographic variables.

The thirty-seven input variables were factor analyzed and the factor matrix rotated via Kaiser's varimax method. The two highest loading non-achievement variables in each of the resulting seven factors were selected for use in the regression analyses, provided that the loading was at least .40. Two mobility variables, percent transfer-in and percent transfer-out, were selected because of their use in previous research. The final list of input variables consisted then of the six subtests of the third grade Metropolitan Achievement Test (Word Knowledge, Word Analysis, Reading Comprehension, Mathematics Computation, Mathematics Concepts, and Problem Solving) and fifteen mobility and demographic variables.

The fifteen mobility and demographic variables included in the regression equations have been given the following operational definitions by the School Information System:

> Percent of Black Pupils (Black Membership): Estimated percent of black pupils. Percent obtained by taking estimated number and dividing by school's average daily membership for a typical day.

Percent Transfers-In: Percent of pupils entering a school from another Cincinnati school after permanent enrollment day. Percent obtained by taking number and dividing by gross membership.

Percent Transfers-Out: Percent of pupils leaving a school for another Cincinnati school after permanent enrollment day. Percent obtained by taking number and dividing by gross membership.

Percent Leaving School: Percent of pupils leaving school after permanent enrollment day for all other reasons than going to another Cincinnati school or moving out of the Cincinnati school district (i.e., going to work, illness, dropouts, etc.). Percent obtained by taking number and dividing by gross membership (for elementary).

Percent First Grade Promotions: Percent of boys and girls promoted from the first grade to the second grade. The percent represents the quotient when the number of boys and girls promoted to the second grade is divided by the number of boys and girls in membership at the end of the school year.

Average Years Teaching--System: Average number of years teachers have taught in the Cincinnati Public Schools. Number of years divided by number of teachers on staff yields this average. Data obtained from Staff Analysis Report, Staff Personnel Branch.

Average Age of Staff: Average age in years of staff members in each school. Calculated by totaling age of each staff member and dividing by number on staff.

Pupil/Teacher Ratio, Total School: Average number of pupils in kindergarten through sixth grade divided by regular teachers in same grades.

Black Percent in Community: Percentage of black persons residing in the community. Data obtained from 1970 Census Book.

Average Income of Families: This figure is the average income for a "typical" family in a school attendance area. This figure is obtained from the 1970 Census Book.

Percentage of Families Below Poverty Level: Percent of families below the defined poverty level. Percent obtained by taking number of families below poverty level and dividing by number of families in school attendance area.

Percent of Registrants Voting: Percent of persons in a school attendance area who actually voted. Percent found by taking number voting and dividing by number of registrants.

Percentage Employed: Total percent employed in a school area as either professional, technical, kindered workers, laborers, etc. Percent of people employed obtained by taking number of persons employed and dividing this number by total number of persons 16 years of age and over.

Absence Per Employee: Days absent for certificated employees or for itinerant and substitute teachers assigned to a school. Calculated by taking total days of staff absence and dividing by number of staff.

Percent of Capacity Being Used: This figure represents how many students are actually occupying the building. This percentage is obtained by taking the capacity of the building and dividing by average daily membership.

For each regression analysis, the dependent variable was the school unit mean on the Reading Comprehension subtest of the Metropolitan Achievement Test battery. A step-wise multiple regression procedure was used for these analyses. The input measure that had the highest correlation with the output measure was selected first. The input measure which added most to the multiple correlation, after control for the first measure, was selected next for inclusion. This process was repeated by adding input measures to the equation until the squared multiple correlation increased by less than .01.

Residuals were computed as the simple difference between the observed school unit value on the dependent variable, and the predicted value obtained from a particular multiple regression equation.

Results

Achievement test scores were available on 7161 students in the sixty-seven elementary schools at either the third grade (April 1971) or the sixth grade (October 1973). Table 1 summarizes the composition of both the matched group of students, the group for which both third and sixth grade

achievement test scores available, and the unmatched group of students, the group containing all the students tested at the third and sixth grades.

(INSERT TABLE 1)

It will be seen in Table 1 that 1784 students, or some 29 percent of the third grade population did not appear for testing in the sixth grade. Most of these students moved out of the system; however, some of the missing students failed a grade between the testing dates, or missed the sixth grade testing for some other reason. Seventeen percent of the total sixth grade population, 893 students, were not tested with the third grade population. The majority of these students moved into the system from some other school district between the testing dates; however, some of them were failures from the previous cohort or students who missed the third grade testing for some other reason.

The mean population for individual schools using the matched group was 67. Populations across all the schools for the matched group of students ranged from 20 to 162. The mean population for individual schools using the unmatched students was 80. Populations across all the schools with the unmatched students group ranged from 25 to 192.

The intercorrelations among the school means on the input and criterion achievement variables for both the matched group (Method II) and the unmatched group (Method III) are given in Table 2. The simple correlations between the school means on the six third grade achievement variables with the school means on the criterion variable, sixth grade reading comprehension, range from .89 to .94 with the matched group, and from .87 to .93 with the unmatched group. In both cases the correlations of third grade reading

comprehension means with sixth grade reading comprehension is slightly higher than the correlations of the other input achievement variables with the sixth grade reading comprehension means.

(INSERT TABLE 2)

The intercorrelations among the achievement test input variables for the matched group and the unmatched group are uniformly high. Intercorrelations for the matched group range from .88 to .97, with a median correlation of .93. Intercorrelations for the unmatched group range from .85 to .98 with a median correlation of .92.

The correlations of the various input achievement test means for the matched group with the corresponding means for the unmatched group range from .96 to .98. The correlations between the means on sixth grade reading comprehension test, between the matched and unmatched group, was .99.

The correlations between each of the non-cognitive variables and the sixth grade reading comprehension scores are presented in Table 3. The differences in the correlations of individual variables with the matched group sixth grade reading comprehension scores and the unmatched group reading comprehension scores are small.

(INSERT TABLE 3)

The multiple correlations between the set of third grade input measures and the sixth grade reading comprehension scores for both Method II and Method III are given in Table 4. In the cases of both Method II and Method III; non-achievement variables contributed

significantly to the multiple regression prediction of school outputs, to the extent that they displaced input achievement variables in the stepwise multiple regression process, and to the extent that they contributed at least .01 to t' tiple correlations. However, the simple correlation (best individual predictor, previous eading comprehension means, was so high that the proportion of additional variability accounted for by non-achievement variables is relatively low.

(INSERT TABLE 4)

However, the total multiple correlation of the non-achievement variables alone with the school means sixth grade reading output was not only in the same order of magnitude as the prediction made by previous achievement, but was not significantly worse than those predictions for either Method II or Method III.

The magnitudes of the multiple correlations are not an appropriate basis for comparing any one method or set of input variables with any other method or set of input variables for the purpose of deriving residuals as a measure of school performance. For the purpose of this study, the relevant basis for comparison is to be found, rather, in the degree to which the deviations from the regression surfaces are similar or dissimilar across methods and sets of input variables.

The intercorrelations between Method II and Method III range from .75 to .83. The residuals derived from Method II and those derived from Method III are comparable over the three different sets of input variables. The intercorrelations of the residuals from both Method II and Method III over the three sets of input variables are given in Table 5.

(INSERT TABLE 5)

The correlations of all the residuals from the regression equations containing non-achievement variables with the residual from the matched group regression equations containing previous achievement variables alone (Method II) indicate that residuals for the matched group based on both achievement and non-achievement variables, and residuals for the unmatched group based on achievement variables alone, are highly comparable ($r \approx .80$). Residuals based on both achievement and non-achievement variables for the unmatched group are moderately correlated with those using Dyer's Method II ($r \approx .64$). Residuals based on non-achievement variables alone are not significantly correlated with those based on Dyer's Method II ($r \approx .20$).

In general, the intercorrelation of the residuals from the four sets of input variables within Methods II and III are in the same order of magnitude.

Discussion

The results reported in this study were obtained on a population different in certain respects from the population used by Dyer (1969). The basic unit of study in this population was the elementary school, whereas, the basic unit under study by Dyer was the school system.

Dyer reported much larger changes between the input and output populations than those found in the present study. In Dyer's study, the matched group represented 56 percent of the total fifth grade population and 70 percent of the sixth grade population. In the present study, the group represented 72 percent of the third grade population and 83 percent

of the sixth grade population. The smaller differences between the matched and unmatched groups could easily account for the higher intercorrelation between residuals across methods found in the present study.

In the case of both the matched group (Method II) and the unmatched group (Method III), non-achievement variables contributed significantly to the multiple regression prediction of school outputs. The simple correlations of the input achievement predictors, however, are so high that the proportion of additional variability accounted for by the non-achievement variables is relatively low.

Superficially, it would seem that mobility and demographic variables contribute a modest but relatively insignificant amount to the prediction of school outputs when they are entered in competition with measures of previous achievement. This interpretation could be sustained whether the data were based on the unmatched student population (Method III), or on the smaller matched student population (Method II).

If the significant question for educational practice is reduced to the relative comparability of residuals based on the three sets of input variables across the matched and unmatched groups, then the answer is fairly simple based on the particular population of schools. School residuals based on the unmatched student data are highly comparable to residuals based on matched student data, when the the inputs are restricted to achievement variables alone ($r \approx .8$). Dyer (1969), using the school system data, found much smaller correlations between matched population and unmatched population residuals ($r \approx .32$), and concluded that the residuals based on matched population are interestingly superior to those based on unmatched population.

The significance of the comparability of residuals based on matched population and unmatched population is essentially a cost effectiveness question. If a school system can arrive at essentially the same decision on the effectiveness of individual schools using unmatched school data, then the clerical and data processing task of building a matched longitudinal data base would be unnecessary. Correlations of residuals in the order of .8 are usually interpreted as evidence of impressive comparability by researchers. However, the 30 to 40 percent of idiosynchratic variability, not common to these two estimates of school effectiveness, would be critical to administrators basing personnel and program decisions on residuals as indicators of school effectiveness.

When non-achievement variables were added to Method III, the resulting residuals were less highly correlated with those resulting from Method II using achievement variables than those which contained only achievement variables. There is no indication that the addition of either mobility or school background variables served to make unmatched residuals more similar to matched residuals based on achievement variables alone.

Within both Method II and Method III the addition of non-achievement variables had similar effects on the comparability of the resulting residuals. In each case the residuals based on achievement variables alone were strongly correlated with those based on combinations of achievement and non-achievement variables ($r \approx .8$), but unrelated to those based on non-achievement variables alone ($r \approx .2$). It appears that school performance indices, based on a combination of achievement and non-achievement variables occupy a position somewhere in between the independent predictions supplied by either of these sets of variables alone. The lack of relationship between residuals based on achievement and non-achievement variables was reported without supporting evidence by O'Connor (1972).

The decision to use non-achievement input variables in the production of residuals as measures of school effectiveness cannot be made on the basis of empirical studies. The existence of different sets of school unit residuals based on the use of different sets of input variables suggests that specification of the input variables is perhaps the crucial decision for the interpreter of residual gain score.

The non-achievement variables used in the present study were selected pragmatically from the kinds of data which exist in most school systems. It is highly doubtful that these existing data sources exhaust the range of variables which affect the performance of schools.

As the range of potential input variables increases, the question of the validity of the resulting residuals becomes more important, especially if the residuals are used to make decisions about the accountability of administrators or other staff.

Future studies in this area will have to move from purely methodological solutions to the evaluative problems involved, to research on the causal models which underlie these regression solutions.

TABLE 1

Summary of Student Data Available
On Both Matched and Unmatched Groups

Student Sample	Grade:	Method II Matched Group 3	Method II Matched Group 6	Method III Unmatched Group 3	Method III Unmatched Group 6
Attended the same school		3378	3378	3378	3378
Moved within the system		1106	1106	1106	1106
In third grade, not sixth		----	----	1784	----
In sixth grade, not third		----	----	----	873
Total Student Sample		4484	4484	6268	5377

TABLE 2

Intercorrelations Between School Means on the Input and Criterion Achievement Variables for the Matched and Unmatched Groups

	1	2	3	4	5	6	cr	7	8	9	10	11	12	cr
Matched Third Grade:														
1) Reading Comprehension	1.0													
2) Word Knowledge	.97	1.0												
3) Word Analysis	.95	.96	1.0											
4) Math Computation	.92	.90	.88	1.0										
5) Math Concepts	.95	.94	.93	.93	1.0									
6) Problem Solving	.94	.93	.91	.93	.95	1.0								
cr) 6th gr. Reading Comprehension	.94	.92	.89	.89	.92	.92	1.0							
Unmatched Third Grade:														
7) Reading Comprehension	.98	.97	.95	.91	.92	.92	.93	1.0						
8) Word Knowledge	.96	.93	.95	.89	.92	.91	.91	.98	1.0					
9) Word Analysis	.91	.93	.96	.84	.86	.88	.87	.94	.95	1.0				
10) Math Computation	.92	.89	.87	.98	.91	.92	.89	.92	.89	.85	1.0			
11) Math Concepts	.95	.93	.92	.92	.99	.94	.91	.94	.93	.88	.92	1.0		
12) Problem Solving	.95	.93	.92	.92	.96	.97	.91	.95	.93	.88	.93	.96	1.0	
cr) 6th gr. Reading Comprehension	.94	.92	.88	.89	.91	.92	.99	.93	.90	.87	.89	.91	.91	1.0

TABLE 3

Correlations Between Non-Cognitive Variables and Sixth Grade Reading Comprehension for Matched and Unmatched Groups

Variable	Matched Group	Unmatched Group
13. Percent Black Membership	-.69	-.71
14. Percent Transfers In	-.78	-.79
15. Percent Transfers Out	-.68	-.70
16. Percent Leaving School	-.21	-.22
17. Percent First-Grade Promotions	.37	.40
18. Average Years Teaching	.32	.29
19. Average Age of Staff	.28	.25
20. Pupil/Teacher Ratio	.47	.50
21. Black Percentage in Community	-.68	-.70
22. Average Income	.81	.82
23. Percent Below Poverty Level	-.76	-.79
24. Percent Registrants Voting	.84	.86
25. Percent Employed	.49	.52
26. Absence Per Employee	-.10	-.10
27. Percent Capacity Being Used	.25	.28

TABLE 4

Multiple Correlations and Input Measures Included in the Multiple Regression Equation, In Order of Entry, For Each Method, Under Each Condition

Input	Input Measures Included	Method II Matched Population	Input Measures Inlcuded	Method III Unmatched Population
Achievement inputs alone	1,6*	.946	7,11	.934
Achievement, mobility, and school background inputs	1,24,16, 27	.962	7,24,22, 13,11,25	.958
Mobility and school background inputs alone	24,22,27, 14,15,25	.922	24,22,14, 15,13,21, 25,27	.950

TABLE 5

Intercorrelations of the Residuals of School System Means From the Regression Surfaces for Both Matched and Unmatched Groups Under Four Sets of Input Variables

Variables	1	2	3	4	5	6	7
1. Achievement (M)*	1.0						
2. Achievement and non-achievement (M)	.81	1.0					
3. Non-achievement (M)	.21	.55	1.0				
4. Achievement (U)**	.80	.69	.25	1.0			
5. Achievement and non-achievement (U)	.64	.75	.57	.80	1.0		
6. Non-achievement (U)	.18	.44	.83	.31	.63	.65	1.0

*Matched Group
**Unmatched Group

REFERENCES

Coleman, J., et. al. Equality of educational opportunity. Washington, D.C.: U.S. Government Printing Office, 1966.

Convey, J. A validation of three models for producing school effectiveness indices. A paper presented at the annual meeting of the American Educational Research Association. Washington, 1975.

Dyer, H. The Pennsylvania plan. Science Foundation, V. 50, pp. 242-248, February 1966.

Dyer, H., Linn, R. and Patton, M. A comparison of four methods of obtaining discrepancy measures based on observed and predicted school system means on achievement tests. American Education Research Journal, V. 6, pp. 591-606, Fall 1969.

Forsythe, R. Some empirical results related to the stability of performance indicators in Dyer's student change model of an educational system. Journal of Educational Measurement, V. 10, pp. 7-12, Spring 1973.

Gastright, J. Some empirical evidence on the stability of discrepancy measures based on observed and predicted school means on achievement tests. A paper presented at the annual meeting of the American Educational Research Association, Chicago, 1974.

Jencks, C., Smith, M., et. al. Inequality: a reassessment of the effect of family and schooling in America. New York: Basic Books, 1972.

O'Connor, E. Extending classical test theory to the measurement of change. Review of Educational Research, V. 42, pp. 73-97, Winter, 1972.

Shaycoft, M. The high school years: growth in cognitive skills, Interim Report 3. Palo Alto, California: American Institute for Research, 1969.