DOCUMENT RESUME

ED 078 055 TM 002 871

AUTHOR	Marco, Gary L.
TITLE	A Comparison of Selected School Effectiveness Measures Based on Longitudinal Data.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RB-73-70
PUB DATE	Mar 73
NOTE	46p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Achievement Tests; *Comparative Analysis; *Effective Teaching; *Elementary Schools; *Evaluation Methods; Grade 3; *Longitudinal Studies; Prediction; Sampling; Test Results; Urban Areas
IDENTIFIERS	*Metropolitan Primary II Achievement Test

ABSTRACT

The purpose of this study was to compare five methods of computing school effectiveness indices (SEIs) from longitudinal data. The methods were within-schcol regression, within-school regression corrected for the unreliability of measurement, mean difference scores, average individual residual scores, and school residual scores. The sample consisted of 3,769 third-graders from 70 elementary schools in the Midwest. The raw data consisted of Total Reading scores from the Metropolitan Primary II Achievement Test administered in fall 1970 and spring 1971. While the various school effectiveness indices differed from one another and in their correlations with other variables, little evidence could be found for the lack of validity of any school effectiveness index. Further, all of the school effectiveness indices were highly stable across samples, except for the indices for initially high-scoring students. Finally, predictions from nonlongitudinal data furnished reasonable estimates of school effectiveness as measured by one of the indices. (Author/CK)

RB-73-20

RESEARCH BULLETIN

A COMPARISON OF SELECTED SCHOOL EFFECTIVENESS MEASURES BASED ON LONGITUDINAL DATA

Gary L. Marco

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
March 1973

A COMPARISON OF SELECTED SCHOOL EFFECTIVENESS

MEASURES BASED ON LONGITUDINAL DATA

Gary L. Marco

Educational Testing Service

Abstract

The purpose of this study was to compare five methods of computing school effectiveness indices (SEIs) from longitudinal data. The five methods were within-school regression, within-school regression corrected for the unreliability of measurement, mean difference scores, average individual residual scores (based on the regression of student output scores on student input scores), and school residual scores (based on the regression of school mean output scores on school mean input scores). The sample consisted of 3,769 third-graders from 70 elementary schools in the Midwest. The raw data consisted of Total Reading scores from the Metropolitan Primary II Achievement Test administered in Fall 1970 and Spring 1971.

While the various school effectiveness indices differed from one another and in their correlations with other variables, little evidence could be found for the lack of validity of any school effectiveness index. Further, all of the school effectiveness indices were highly stable across samples, except for the school effectiveness indices for initially high-scoring students. Finally, predictions from nonlongitudinal data furnished reasonable estimates of school effectiveness as measured by one of the school effectiveness indices.

The methods should be tried out at other grade levels. Further, the stabilities of the various school effectiveness indices across years should be studied.

A COMPARISON OF SELECTED SCHOOL EFFECTIVENESS MEASURES BASED ON LONGITUDINAL DATA[1]

Gary L. Marco

Educational Testing Service

With the recent emphasis in education upon program budgeting and cost effectiveness has come a renewed interest in school system evaluation. However, how school effectiveness should be estimated is unclear. The purpose of this study is to compare selected methods of estimating school effectiveness from longitudinal data.

Various techniques have been suggested to generate school effectiveness indices. Indices commonly used are the average performance of students in a particular grade in the school and the difference between the performance of students in the school and the performance of a national norm group. Although these two methods have been widely used, they have a fatal flaw: neither takes into account the differences in initial status.

In some studies partial control over differing input levels has been achieved by holding socioeconomic status (SES) constant. Schools serving students from low SES families have been compared with one another, as have schools serving students from more advantaged families. The school effectiveness index in such a case is the deviation of performance from the average of the schools serving like children. This index is often employed with data collected at one point in time for a given grade level, such as statewide testing program data. Ability scores have sometimes been partialed out of achievement scores in an attempt to control for initial differences. In this case, the difference between the actual performance and the predicted performance has been used as a measure of school effectiveness. Unfortunately, the distinction between ability and achievement is unclear operationally, so that partialing out ability also partials out some of the valid school variance. In other studies cross-sectional data have been used to estimate school effectiveness indices. These data are useful for estimating school effectiveness only if it is assumed that students in the lower grade are now performing on the average at the same level as students in the higher grade did at the lower grade level. The difference in the means of the two groups has been used as a measure of effectiveness.

Longitudinal data have been recognized as the sine qua non of good evaluation in nonexperimental settings (see Dyer, Linn, & Patton, 1969; Hilton & Patrick, 1970). Longitudinal data may be available at the school level (for example, third grade arithmetic mean and sixth grade arithmetic mean three years later) or at the student level. Unless the student group enrolled in a lower grade has remained intact over the interim period, the school data will be based on a group that is somewhat different from the group of students that was present at both data collection points. To distinguish these "unmatched" groups from groups that are composed of the same students, the former has been called an "unmatched-longitudinal sample" and the latter a "matched-longitudinal sample" (Dyer et al., 1969).

Dyer, Linn, and Patton compared school effectiveness indices based on a matched-longitudinal sample, an unmatched-longitudinal sample, and a cross-sectional sample, and concluded:

> Although it seems apparent that the use of discrepancy measures raises a great many problems needing further research, it is also apparent, from the present study, that such measures when based on matched-longitudinal

> samples are the ones most likely to provide valid measures of system effectiveness [1969, p. 605].

There are a number of methods of generating school effectiveness indices that could be used with longitudinal data that were not used in the Dyer study. The purposes of the present study were (a) to compare the school effectiveness indices generated by the selected methods, (b) to estimate the stability of the school effectiveness indices across samples, and (c) to assess the adequacy of using nonlongitudinal data for predicting school effectiveness indices obtained from longitudinal data. Three sub-studies were conducted to accomplish these purposes.

General Procedures

The general procedures for the study are outlined in this section. Procedures specific to the sub-studies are discussed in the three following sections.

Sample. The schools in the sample consisted of 70 elementary schools that participated in a 1970-71 ESEA Title I statewide evaluation study conducted in the Midwest. The students in the sample were those third-graders who took a pretest in the Fall (November primarily) of 1970 and a posttest in the Spring (May primarily) of 1971. Only those students tested in both the Fall and the Spring were included in the study sample. A total of 4,778 students were tested at least once; 3,769 students (79%) were tested both times. The sample sizes for the schools in the study ranged from 17 to 152.

Instruments. Forms F and G of the Primary II Reading Test of the Metropolitan Achievement Tests were used for the study. The Primary II Reading Test is appropriate for second- and third-graders. Since the third-graders in the sample, being students in Title I and comparison schools, were assumed to be lower achievers than the average third-grader, the Primary II Reading Test was administered to ensure that the test material would not be too difficult for the students. Total Reading standard scores from the two administrations were used as the data base for the study.

The appropriateness of the Primary II Reading Test for the sample is indicated by the below-average pretest means of the sample. The pretest mean reading score for the 3,769 students was 51.82, which corresponds to a grade equivalent score of 2.6. Thus, the study sample was on the average about six months behind the norm for students in the second month of their third year.

Variables. Information on a number of variables was used in the study. These are listed in Table 1. Records of the State Department of Education

----

Insert Table 1 about here

----

furnished information on Variables 1-9. A questionnaire about Title I reading programs yielded data on Variables 10 and 11, while Variables 12-15 were obtained directly from the Metropolitan Achievement Test.

Since schools tested on different dates, the number of days between the pretest and the posttest was not the same from school to school. The number of weekdays between testings ranged from 119 days to 160 days. Variable 16, Weekdays between Pretest and Posttest, was derived from the testing dates.

The staff of each school was asked to identify those students participating in an ESEA Title I reading program. This information, coupled with the number of students taking both tests, defined Variable 17, Percent of Students Participating in Title I Reading.

Student stability also varies from school to school. To obtain a "stability" index for the school, the total number of students who took both tests was divided by the total number of students tested, the assumption being that those students not tested twice transferred either in or out during the year. This variable was Percent of Students Taking Both Tests (Variable 18).

Variables 19-30 were those associated with the various school effectiveness indices that were derived. Variables 19-21, 23-26, 28, and 30 were of primary interest in Sub-studies I and II. Variables 1-18, 22, 27, and 29 were useful in interpreting the school effectiveness indices in Sub-study I. Variables 1-9 and Variable 18 were used as predictors in Sub-study III.

Methods of estimating school effectiveness. Five methods were used to compute school effectiveness indices. These were as follows:

1. Within-school regression. For each school the regression line describing the relationship between individual student pretest and posttest scores was computed and posttest values were estimated at reference points for low-, middle-, and high-scoring students. That is, for a given reference point, X_0 ,

$$SEI = AX_0 + \overline{Y} - A\overline{X}$$
$$= \overline{Y} - A(\overline{X} - X_0)$$

   where A is the least squares estimate of the within-school slope; and $\overline{Y}$ and $\overline{X}$ are the school means on the posttest and pretest, respectively.

2. Corrected within-school regression. Same as 1, except the slope and intercept were corrected for the unreliability of the pretest

measure on the basis of data reported by the test publisher. Symbolically,

$$\text{SEI} = \frac{A}{R_{XX}} X_0 + \bar{Y} - \frac{A}{R_{XX}} \bar{X}$$

$$= \bar{Y} - \frac{A}{R_{XX}} (\bar{X} - X_0)$$

where A, X_0, $\bar{Y}$, and $\bar{X}$ are defined as in Method 1; and R_{XX} is the pretest reliability estimate for the school.

3. Mean difference scores. For each school the mean difference score (posttest score minus pretest score) was computed; thus,

$$\text{SEI} = \bar{Y} - \bar{X} \ .$$

4. Individual residual scores. Individual student posttest scores were regressed on individual student pretest scores for the total sample of students across schools an n individual residual scores calculated for each school. In this case,

$$\text{SEI} = \frac{1}{N} \sum_i [Y_i - (BX_i + \bar{\bar{Y}} - B\bar{\bar{X}})]$$

$$= \bar{Y} - [\bar{\bar{Y}} - B(\bar{\bar{X}} - \bar{X})]$$

where N is the number of students in the school taking both tests; Y_i and X_i are the posttest and pretest scores, respectively, for individual i ; B is the least squares estimate of the slope for the students across all schools; $\bar{\bar{Y}}$ and $\bar{\bar{X}}$ are the grand posttest and pretest means, respectively; and $\bar{Y}$ and $\bar{X}$ are the posttest and pretest means for the school.

5. School residual scores (pretest model). School posttest means were regressed on school pretest means and school residual scores calculated. This is the one-predictor-variable equivalent of the method suggested by Dyer (1971) as a measure of school effectiveness. Here

$$\begin{aligned} SEI &= \overline{Y} - (C\overline{X} + \overline{\overline{Y}}' - C\overline{\overline{X}}') \\ &= \overline{Y} - [\overline{\overline{Y}}' - C(\overline{\overline{X}}' - \overline{X})] \end{aligned}$$

where $\overline{Y}$ and $\overline{X}$ are posttest and pretest means, as previously defined; C is the least squares estimate of the regression slope of the school posttest means on the school pretest means; and $\overline{\overline{Y}}'$ and $\overline{\overline{X}}'$ are the unweighted averages of the school posttest and pretest means, respectively, across all schools.

It should be pointed out that the school effectiveness indices derived from the five methods are not comparable in the absolute sense. The relative positions of the schools on the various school effectiveness indices may be compared, however.

Most of the methods are straightforward computationally. However, an elaboration of the first two models is in order, since these models differ from those that have been used to estimate school effectiveness.

The first two models are similar to analysis of covariance except that no assumption is made about the equality of slopes from school to school. This assumption about slopes is particularly restrictive when one deals with existing groups, such as the students in schools. Model 1 relates the observed posttest score to the observed pretest score. This procedure is like the one Rock, Baird, and Linn (1972) used to estimate college effectiveness.

Often one is interested in comparing treatment groups not formed at random, as in the case where schools are being compared. Cronbach and Furby (1970) indicated that the findings of such a study can usefully be summarized by calculating the within-group regression equations relating true status on the posttest to true status on independent variables. Model 2 does essentially what Cronbach and Furby recommended; the slope and intercept of the regression line relating the posttest to the pretest is corrected for the unreliability of the pretest measure.

In each of the first two models, it is assumed for purposes of the present study that a straight line best describes the relationship between the pretest scores and the posttest scores for a given school. It is not assumed that the regression lines are the same from school to school, for a school may be more effective for one type of student than another. In Figure 1 it may be noted that School A appears to be the most effective for

----

Insert Figure 1 about here

----

high-scoring students, while School B appears to be the most effective for low-scoring students. It is obvious in this case that a single school effectiveness index is a misleading index, since it does not indicate the fact that schools are differentially effective for students of differing abilities. The school effectiveness index depends upon which pretest score is selected as the reference point.

When only three schools are considered, as in Figure 1, a reference point need not be selected; the graph itself is an adequate description of school effectiveness. However, as the number of schools increases, a graph of the lines becomes very messy; and it is necessary to resort to a nongraphical procedure for describing school effectiveness. If the slopes of the regression lines were equal, then one could choose any arbitrary point on the pretest score scale and compare predicted posttest scores. The schools would maintain the same ordering, no matter which points were chosen. However, if the slopes were not equal, as is likely to be the case if existing groups are studied, the selection of a reference point is crucial.

For purposes of the present study, it was decided to select reference points to represent low-scoring, middle-scoring, and high-scoring students and to compute a school effectiveness index for each group, as shown in Figure 1. The points selected were the mean pretest scores across all schools and those points located one standard deviation above and below the mean; namely, 40.8892, 51.8201, and 62.7510. The school effectiveness indices are estimates of the mean posttest scores for individuals having these fixed pretest scores. While one could subtract the fixed point from the estimated mean to obtain a "growth" school effectiveness index, the estimated mean works just as well and is used in this study as the school effectiveness index.

In Model 2, corrections for the unreliability of measurement were made. It is well known that test unreliability results in an underestimate of the slope of Y on X (see, for example, Snedecor & Cochran, 1967). The sample regression coefficient, as Snedecor and Cochran indicate, provides an unbiased estimate of $A'(R_{XX})$, where A' is the true slope and R_{XX} is the reliability of the predictor variable for group X. Thus, A' can be estimated by dividing A, the observed slope, by the reliability of the test.

The problem of bias due to measurement error is most serious when the posttest scores are estimated from regression lines determined on groups that have widely disparate pretest means. Suppose that two groups had the same observed slopes and intercepts, but differed only in mean performance on the pretest and posttest, as illustrated in Figure 2. For any selected

----

Insert Figure 2 about here

----

reference point on X , the estimated value of Y would be exactly the same for the two groups. Suppose, however, that the slopes (and intercepts) were corrected for measurement error. They would then appear somewhat as shown by the dotted lines. The expected value of the group with the lower mean would be higher for any reference point, say X_0 . In this case, the estimated values computed on the basis of the observed slopes and intercepts would be biased against the lower scoring group--a phenomenon that will obtain whenever the slope is positive. The adjustment of the slopes requires the use of the test reliability.

In this study test reliabilities for each school were not available; they had to be derived. If it is assumed that the error variances of two groups are the same, then the following formula (Gulliksen, 1950, p. 111) can be used for estimating the reliability of test scores for Group 2 from the reliability of Group 1 scores:

$$R_{22} = 1 - \frac{S_1^2}{S_2^2} (1 - R_{11})$$

where S_1^2 is the variance of Group 1 on the test,

R_{11} is the reliability of the test for Group 1,

S_2^2 is the variance of Group 2 on the test, and

R_{22} is the reliability of the test for Group 2.

But $S_1^2(1 - R_{11})$ is the variance error of measurement. Thus, one can estimate the reliability of the pretest for a particular school if one knows the Total Reading variance error of measurement for the Metropolitan standardization sample and the variance of Total Reading scores for the school.

The standard error of measurement for the standardization sample on Total Reading was 1.9 (Durost, Bixler, Wrightstone, Prescott, & Balow, 1971); the variance error of measurement was thus $(1.9)^2$ or 3.61. Hence, the pretest reliability for a given school was estimated from the formula:

$$R_{xx} = 1 - \frac{3.61}{S_x^2} .$$

The estimated true slopes and intercepts for a given school were computed as follows:

$$A' = \frac{A}{R_{xx}}$$

$$B' = \overline{Y} - A'\overline{X}$$

where A' is the corrected slope,

A is the observed slope,

R_{xx} is the estimated pretest reliability,

B' is the corrected intercept,

$\overline{Y}$ is the posttest mean, and

$\overline{X}$ is the pretest mean.

The corrected school effectiveness index for a given reference point, X_0, was computed by the formula $A'X_0 + B'$.

Sub-study I: Comparison of School Effectiveness Indices

Procedures and Results

The first sub-study involved a comparison of the school effectiveness indices generated by the five methods. The school effectiveness indices were computed according to the methods already outlined for each of the 70 schools in the sample. Regression coefficients and other descriptive information for the schools that had the highest and lowest school effectiveness indices, as estimated from the corrected within-school regression lines, are reported in Table 2. The regression coefficients for the two other

----

Insert Table 2 about here

----

models using regression lines to generate the school effectiveness indices are shown in Table 3. It may be noted that the slope for the School Residual

----

Insert Table 3 about here

(Pretest) method is very close to one. Thus, school effectiveness indices generated using this model (observed mean minus estimated mean) will necessarily be highly correlated with Mean Difference Scores.

Intercorrelations among the school effectiveness indices derived from the five methods are reported in Table 4. The intercorrelations among the

----

Insert Table 4 about here

----

nine school effectiveness indices were factor analyzed by the Minres method (Harman, 1967). The residual correlations were negligible after three factors had been extracted. The three derived factors rotated according to the normalized varimax criterion are shown in Table 5.

Insert Table 5 about here

The other 21 variables identified in the section on Variables were correlated with the school effectiveness indices to provide the basis for comparing the methods. These correlations are given in Table 6.

Insert Table 6 about here

In Tables 4 and 6 a correlation of ±.235 is significantly different from zero at $\alpha = .05$. A difference of $\pm.15$ in the correlations of any two school effectiveness indices with a third variable is significant at $\alpha \leq .05$ if the correlation between the school effectiveness indices is at least .80. (This difference is conservative, for it assumes a multiple correlation of zero between a weighted combination of the school effectiveness indices and the third variable. See Dubois, 1965, p. 349, for the exact test.)

<u>Discussion: Direct Comparisons of the Estimates</u>

With respect to the correlations shown in Table 4, it may be noted, first of all, that the corrected school effectiveness indices correlated nearly perfectly with their corresponding uncorrected school effectiveness indices (19 vs. 23, 20 vs. 24, 21 vs. 25). Thus, in this study the correction for the unreliability of the pretest made little difference. This effect is not surprising in view of the fact that the Total Reading score on the Metropolitan Reading Test had a reliability of .97 for the national norm group. In this study, then, Variables 23, 24, and 25 were virtually interchangeable with Variables 19, 20, and 21, respectively.

Secondly, the Mean Difference Scores (Variable 28) correlated nearly perfectly (*r* = .996) with the School Residual (Pretest) School Effectiveness Indices (Variable 30). This result, too, is not surprising in view of the fact that the slope for the School Residual (Pretest) School Effectiveness Index was .96 (see Table 3). If the slope for the regression of school mean output or school mean input were 1.00, the two variables would have been perfectly correlated.

Thirdly, the Individual Residual School Effectiveness Indices were more highly correlated with the School Residual (Pretest) School Effectiveness Indices (*r* = .96) than any other type of school effectiveness index. Both of these methods utilize as school effectiveness indices deviations about the regression line of a reference group. A school's school effectiveness index is the difference between the observed school posttest mean and the predicted school posttest mean. Although the regression coefficients for the two models were different (see Table 3), apparently the higher intercept for the Individual Residuals compensated enough for the lower slope to yield predicted school posttest means that were similar to those computed from the School Residual regression coefficients.

Fourthly, assuming the validity of the Within-School Regression Corrected School Effectiveness Index, the schools were differentially effective for low-, middle-, and high-scoring students. The correlation of the school effectiveness indices for low-scoring students (Variable 23) with the school effectiveness indices for middle-scoring students (Variable 24) was .79, but the correlation of Variable 23 with the school effectiveness indices for high-scoring students (Variable 25) was only .25. Thus, the rank ordering of the schools changed substantially with the ability of the students. On the basis of the current data, it must be concluded that in general a single school effectiveness index for a school is not an accurate description of the effectiveness of the school for students at all ability levels.

Finally, of the school effectiveness indices not computed from within-school regression (Variables 26, 28, and 30), the Individual Residual School Effectiveness Indices (Variable 26) correlated highest with the various within-school regression (corrected and uncorrected) school effectiveness indices. The correlation of the Individual Residual School Effectiveness Indices with the corrected school effectiveness indices for middle-scoring students ($r = .95$) was considerably higher than the correlations with the corrected school effectiveness indices for low- and high-scoring students (r's · .77 and .73, respectively). Figure 3 shows the relation between

---

Insert Figure 3 about here

Variables 24 and 26. It may be noted that the discrepancy between the two methods increased as the school effectiveness indices increased. Perhaps the within-school regression lines for "high-scoring" schools deviated more from the total individual regression line than did the lines for "low-scoring" schools. This and other hypotheses should be explored with new samples.

Discussion: Factor Analytic Results

The data in Table 5 indicate that three dimensions are necessary to account for the intercorrelations among the school effectiveness indices and further demonstrate that the various methods yield different estimates of school effectiveness. Factors I and II represent school effectiveness for low-scoring and high-scoring students, respectively. Factor III separates the Mean Difference Scores and the School Residual (Pretest) School Effectiveness Indices from the other types of school effectiveness indices. The Within-School Regression (corrected and uncorrected) School Effectiveness Indices (Middle-Scoring Students) and the Individual Residual School Effectiveness Indices have moderate loadings on all three factors. These factor-analytic results lend support to the claim that one school effectiveness index is insufficient for summarizing school effectiveness and to the claim that Mean Difference Scores and School Residuals provide estimates of school effectiveness that are different from those provided by the other methods.

Discussion: "Meanings" of the Estimators

The correlations of the school effectiveness indices with Variables 1-18, 22, 27, and 29 (see Table 6) provide a basis for interpreting the school effectiveness indices. In terms of their correlations with another variable, the variables were, with few exceptions, ordered (high-to-low or low-to-high) as follows: Variable 24 (or Variable 20), Variable 26, Variable 30, and Variable 28. Differences of $\pm .15$ or more existed between at least one pair of school effectiveness indices in the correlations with Variables 1, 2, 3, 4, 6, 10, 12, 13, 14, and 18. Particularly striking are the differences in the correlations of the school effectiveness indices with Variables 12, 14, 3, 2, and 6.

Several of the school effectiveness indices have relatively high correlations with the Total Reading pretest means. The correlations of the

Within-School Regression Corrected School Effectiveness Indices with the pretest means ranged from .19 to .38. The Individual Residual School Effectiveness Indices had a correlation of .28 with the pretest means. Of course, the School Residual (Pretest) School Effectiveness Indices had no relationship to the pretest means, since they were computed by partialing out the pretest means. The Mean Difference Scores correlated negatively (-.10) with the pretest means.

O'Connor (1972) claimed that a method using school residuals (cf. Variable 30) is preferable to a method using mean individual residuals (cf. Variable 26). His argument was based on the assumption that mean student input and residuals should be uncorrelated. However, the "true" correlation of school effectiveness and the initial input of students, while unknown, might well be positive. Wealthier school districts, which frequently have better facilities and a more experienced, highly trained staff, usually serve higher achieving students. If the schools in such districts were more effective, given equal student input, one would expect a positive correlation between student input and effectiveness and input to be zero.

The possibility must be entertained that the higher school effectiveness indices of schools serving students of higher initial achievement levels were due to the lack of control over relevant variables. Campbell and Erlebacher (1971), in discussing ex post facto evaluations of compensatory education, pointed out: "...The usual procedures of selection, adjustment, and analysis produce systematic biases in the direction of making the compensatory program look deleterious [p. 185]." Barnow (1972), however, demonstrated that under certain conditions ex post facto analysis does not lead to bias.

The negative correlations of Mean Difference Scores with Pretest Mean were not unexpected. It is well known that difference scores tend to be negatively correlated with initial scores (see, for example, Thorndike, 1966). Such a condition would result in a negative correlation between the school effectiveness indices and the Pretest Means. This condition is undesirable in that it produces results biased in favor of initially lower-scoring groups, and it is for this reason that an attempt is made to control initial status. It should be pointed out that sometimes this bias in difference scores can lead to an unbiased measure of effectiveness--when it counterbalances the bias from other sources (Campbell & Erlebacher, 1971).

The correlations of the school effectiveness indices with the Total Reading posttest means are even higher than those with the pretest means. This result is to be expected whenever there are treatment effects (in this case, school effects), for the posttest scores reflect the treatment effects as well as the initial achievement levels.

The correlations of the Within-School Regression (corrected and uncorrected) School Effectiveness Indices (Middle-Scoring Students) and the Individual Residual School Effectiveness Indices with Percent Non-White (Variable 3) are much lower than those for Mean Difference Scores and School Residual (Pretest) School Effectiveness Indices. All of the correlations were negative. The same pattern is true of the correlations with K-12 District Current Operating Expense per Pupil (Variable 2). However, the correlations in this case were positive. Thus, Mean Difference Scores and School Residual (Pretest) School Effectiveness Indices tended to be less correlated (in an absolute sense) with the racial composition of the schools and more correlated with cost per pupil. The correlations with Percent of

Teachers with Five or More Years Experience (Variable 6) were close to zero. Those of the Within-School Regression (corrected and uncorrected) School Effectiveness Indices were slightly positive, while those for Mean Difference Scores and School Residual (Pretest) School Effectiveness Indices were slightly negative.

Summary

This analysis indicates that the school effectiveness indices generated by the five different methods differ and have somewhat different correlation patterns with other variables. However, which of the school effectiveness indices best approximates "true" school effectiveness over one academic year is not known. There is a need for a study of schools of known quality, so that a determination of the validities of the various school effectiveness indices can be made.

Sub-study II· Stability of the Estimates

In the preceding section, differences among the various estimates of school effectiveness were pointed out. One factor that should be considered in choosing a method of estimating school effectiveness is the variance, or mean square errors, of the estimator. A biased estimator may be useful if it is not "too" biased and if the estimates vary little from one sample to another. In this section evidence concerning the stability (reliability) of the various estimates of school effectiveness is presented.

Existing statistical theory could have been used to derive estimates of the reliabilities. However, since such estimates are based on normal distribution theory, which may not apply here, an empirical determination of the reliabilities of the estimators was made. As in the previous sub-study, Total Reading scores were used.

Procedures and Results

The sample from each of the schools in the study was divided into random halves by use of the Tausworthe random number generator for the IBM 360 (Whittlesey, 1968). Then each of the five methods of computing school effectiveness indices were used to estimate school effectiveness for each sample. Ten school effectiveness indices were thus available for each school. (These correspond to Variables 19-21, 23-26, 28, and 30, identified in Table 1.)

The variances of the estimators could not be compared directly because the scales differed from one set of school effectiveness indices to another. Therefore, a scale-free index had to be used. Two such indices were used here, a reliability coefficient and a signal-to-noise ratio.

Reliability coefficients for the school effectiveness indices estimated by each of the five methods were computed by means of analysis of variance. The variation of the school effectiveness indices estimated by a particular method for the two random halves can be divided into among-school variation and within-school variation. The expected mean squares are as follows:

Source of Variation	Expected M.S.
Schools	σ^2 samples + 2 σ^2 schools
Samples within schools	σ^2 samples

As indicated in Winer (1962), the reliability of the mean of two observations is estimated by $\dfrac{\sigma^2 \text{ schools}}{\sigma^2 \text{ schools} + \dfrac{\sigma^2 \text{ samples}}{2}}$. Here, of course, the observations were the school effectiveness indices for the two samples. Each sample was composed of 27 students on the average, but the number varied from school to school. The variation in sample sizes did not enter into the computations here. Thus, reliability estimates were based on an unweighted-means analysis of variance.

For each method a signal-to-noise ratio was also computed. It is simply σ^2 schools divided by the estimated "noise," $\frac{\sigma^2 \text{ samples}}{2}$. This index furnishes another way to look at the stability of an estimator. It is informative because, as Stanley (1971) pointed out, an increase in the ratio is directly related to an increase in number of items (or number of student samples in this case). Thus, by dividing the signal-to-noise ratio of one measure by that of another, one can discover how many samples would have to be used in order to make the reliability of the two measures equal.

Table 7 shows the variance components, signal-to-noise ratios, and reliability coefficients for the school effectiveness indices estimated by the five different methods.

Insert Table 7 about here

Discussion

The Individual Residual School Effectiveness Indices were the most stable across samples, having a reliability coefficient of .85 and a signal-to-noise ratio of 5.61. The Within-School Regression (corrected and uncorrected) School Effectiveness Indices were the least stable, particularly the school effectiveness indices for high-scoring students. The instability of the school effectiveness indices for high- and low-scoring students is to be expected, since under usual conditions data are limited in the extremes.

For instance, under normal distribution theory the variance of the Within-School Regression School Effectiveness Index for a given school is a function of $(X_0 - \overline{X})^2$, where X_0 is the reference point for which the school effectiveness index is to be estimated, and $\overline{X}$ is the mean pretest score for the school (see, for example, Draper & Smith, 1966, p. 22). The variance increases as $(X_0 - \overline{X})^2$ increases, and for most schools would yield larger variances for the school effectiveness indices at the extreme reference point than for the school effectiveness indices at the middle reference point. It should be remembered that indices of school effectiveness for students of differing initial achievement were not computed for three of the methods. Thus, the stability of the Individual Residual School Effectiveness Indices, Mean Difference Scores, and School Residual (Pretest) School Effectiveness Indices for high- and low-scoring students is not known.

The school effectiveness indices for middle-scoring students computed from the within-school regression lines were slightly less stable than the Individual Residual School Effectiveness Indices, Mean Difference Scores, and School Residual (Pretest) School Effectiveness Indices. Comparing the signal-to-noise ratios indicates that it would take 1.4 (5.61/3.92) student samples to make the reliability of Within-School Regression School Effectiveness Indices (Middle-Scoring Students) equal to that of Individual Regression School Effectiveness Indices based on <u>one</u> student sample.

It may be noted that the within-school regression corrected school effectiveness indices were less stable than their corresponding uncorrected school effectiveness indices, presumably because of the error involved in estimating reliability. Since it can be assumed that the corrected school effectiveness indices were less biased, one is forced to choose between a school effectiveness index that is less biased and a school effectiveness index that is more stable.

While the extent of the bias in the various estimates is unknown, if it could be determined that the other methods of computing school effectiveness indices were only slightly more biased than the within-school regression school effectiveness indices, then they might be more useful as measures of school effectiveness because of their greater stability.

In any case, for this sample of schools, all of the school effectiveness indices except those for high-scoring students appear stable enough to warrant their use as measures of school effectiveness. However, with regard to stability the Individual Residual School Effectiveness Indices are the preferred ones.

Sub-study III: Prediction of School Effectiveness Indices

Given that reasonably good estimates of school effectiveness are available from longitudinal data, it may be possible to predict the school effectiveness indices with a reasonable degree of accuracy from nonlongitudinal data that are readily available in many schools. This possibility was investigated in Sub-study III.

Procedures and Results

The Within-School Regression Corrected School Effectiveness Index (Middle-Scoring Students) was selected as the measure of school effectiveness to be predicted from the state variables, and hereafter is referred to as SEI'(M). Because of the correction for the unreliability of the pretests this school effectiveness index was assumed to be less biased than the other school effectiveness indices for estimating <u>overall</u> school effectiveness. The variables used to predict SEI'(M) were Variables 1-9, 18, and two new variables. (See Table 1 for descriptions.) The two new predictor variables were the posttest mean and posttest standard deviation based on <u>all</u> students who took the posttest. These variables are the nonlongitudinal counterparts of Variables 14 and 15, which were based on those students who took <u>both</u> the pretest and the posttest, and are labeled 14' and 15' in the remainder of this section. Such achievement data, based on all students rather than on a longitudinal sample, are often available in schools. Variable 18, Percent of Students Taking Both Tests, was included as a predictor variable even though it was based on test data taken at two points in time. As was indicated previously, this variable is a measure of the stability of the student body. Although a stability index in the form of Variable 18 might not be available in many schools, some index of the stability of the student body would usually be available.

A forward-selection stepwise regression procedure was used to select the predictor variables to be included at each stage. In this process the regression of the variables incorporated into the model in previous stages was examined. Predictors were added until the amount of variance accounted for by any predictor left out of the model was less than .001. The results are given in Table 8. Multiple Rs, standard errors of estimate, and F-tests are reported as well as unstandardized regression coefficients.

Insert Table 8 about here

Discussion

Only three variables, Variables 14', 4, and 2, made a significant contribution to prediction at $\alpha = .05$ (see the bottom line of Table 8). These were Posttest Total Reading Mean, Pupil/Professional Instructional Staff Ratio, and K-12 District Current Operating Expense per Pupil (1969-70). The correlation of these three predictors with SEI'(M) was .79. The equation using all of the variables correlated .83 with SEI'(M) and accounted for 68% of the variance, as opposed to 62% for the three predictors. No validation of these results was attempted; but, if the weights derived from the full equation were used on a new sample, the shrinkage that would probably result makes it desirable to use the three-predictor equation.

It is interesting to observe that the weight for a given variable changed very little as predictors were added. One would have expected that, as a result of the increasing multicollinearity of the predictors, the regression weights would have bounced around. The squared multiple correlation of any one of the predictor variables with the remaining predictor variables is an indication of the collinearity of that variable with the others. In the nine-predictor subset, Variables 8 and 9 had the highest squared multiple correlations (.87) with the other predictors (see Table 8). The highest squared multiple correlation of any one predictor with the remaining predictors in the three-predictor subset was only .06. Despite the increasing multicollinearity, the regression weights remained relatively stable.

It is also interesting to note that Variable 3, Percent of Non-White Students, was not among the nine predictors selected by the stepwise regression program, even though its zero-order correlation with SEI'(M) was -.43 (see Table 6). If it had been entered as a tenth variable, it would have accounted for only 6% of the variance associated with SEI'(M) . Thus, almost all of the SEI'(M) variance accounted for by Variable 3 was also associated with the other predictors.

Assuming the validity of the Within-School Regression School Effectiveness Indices for middle-scoring students, it appears that a reasonable estimate of the effectiveness of a school for a given year can be made from the mean score of a school (from spring data), pupil/professional instructional staff ratio, and current operating expense per pupil. The regression weights obtained in this study would apply only if the same measures were used. However, the standardized regression weights may apply more generally. The standardized weights for Variables 14', 4, 2, respectively, were .78, -.27, and .23, indicating that weighting the standard scores by 3, -1, and 1 would give an approximate indication of school effectiveness. The school effectiveness indices predicted from these three variables are plotted against SEI'(M) in Figure 4.

Insert Figure 4 about here

Conclusions

This study has shown that the five methods of estimating yield school effectiveness indices are highly correlated. However, they are different enough from one another and in their relations with other variables to prevent them from being used interchangeably. The school effectiveness indices for initially low- and high-scoring students appeared to yield unique information and raised doubts about using a simple index to measure school effectiveness. While differences existed among the school effectiveness indices, little evidence could be found for the lack of validity of any school effectiveness index. The methods should be tried out in a situation where the quality of schools is well known, so that a reasonable choice among the estimators can be made.

The study has also shown that, except for the Within-School Regression (corrected and uncorrected) School Effectiveness Index for high scoring students, the various school effectiveness indices were highly stable. However, the stability was measured in terms of the sampling error associated with random halves. A more important kind of stability is the stability of the school effectiveness indices from one year to the next for schools whose physical facilities, staff, student body characteristics, and programs remain basically unchanged. Forsyth (1973) investigated the stability of high-school school effectiveness indices based on the residuals of twelfth grade means from predicted means based on ninth grade data. He found correlations in the .20's for two different longitudinal student samples. The extent to which the various school effectiveness indices included in this study are stable over years is unknown and needs to be studied.

A final conclusion is that over a one year period predictions from nonlongitudinal data furnished reasonable estimates of school effectiveness ($r = .79$), assuming the validity of the Within-School Regression School Effectiveness Index for middle-scoring students. The results here were not validated on a separate sample of schools. While the use of nonlongitudinal data as a substitute for school effectiveness indices based on longitudinal data seems promising, the method should not be used in a practical setting until further evidence is accumulated.

The conclusions from this study are limited in their generalizability in two respects: (a) The study involved an accidental sample of third grade students. The students were in schools that had Title I reading programs and in comparison schools and were somewhat below average in reading achievement. The results may have differed if students of higher ability or at higher grade levels had been involved. (b) The study was limited to longitudinal data collected during one academic year. The methods should be tried out in a situation where a two- or three-year interval exists between pretesting and posttesting and where data on more than one cohort are available.

References

Barnow, B. S. Conditions for the presence or absence of a bias in treatment effect: Some statistical models for Head Start evaluation. Paper presented at a Northwestern University colloquium, Evanston, Illinois, February 1972.

Campbell, D. T., & Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), Compensatory education: A national debate. Vol. III of The disadvantaged child. New York: Brunner/Mazel, 1971.

Cronbach, L. J., & Furby, L. How we should measure "change"--or should we? Psychological Bulletin, 1970, 74, 68-80.

Draper, N. R., & Smith, H. Applied regression analysis. New York: Wiley, 1966.

DuBois, P. H. An introduction to psychological statistics. New York: Harper & Row, 1965.

Durost, W. N., Bixler, H. H., Wrightstone, J. W., Prescott, G. A., & Balow, I. H. Metropolitan achievement tests: Primary II teacher's handbook. New York: Harcourt, Brace & Jovanovich, 1971.

Dyer, H. S. Toward objective criteria of professional accountability in the schools of New York City. Phi Delta Kappan, 1971, 52, 206-211.

Dyer, H. S., Linn, R. L., & Patton, M. J. A comparison of four methods of obtaining discrepancy measures based on observed and predicted school system means on achievement tests. American Educational Research Journal, 1969, 6, 591-605.

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

Gulliksen, H. Theory of mental tests. New York: Wiley, 1950.

Harman, H. Modern factor analysis. (2nd ed., rev.) Chicago: The University of Chicago Press, 1967.

Hilton, T. L., & Patrick, C. Cross-sectional versus longitudinal data: an empirical comparison of mean differences on academic growth. Journal of Educational Measurement, 1970, 7, 15-24.

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

Rock, D. A., Baird, L. L., & Linn, R. L. Interaction between college effects and students' aptitudes. American Educational Research Journal, 1972, 9, 149-161.

Snedecor, G. W., & Cochran, W. G. Statistical methods. (6th ed.) Ames, Iowa: Iowa State University Press, 1967.

Stanley, J. C. Reliability. In R. L. Thorndike (Ed.), Educational Measurement. (2nd ed.) Washington, D.C.: American Council on Education, 1971.

Thorndike, R. L. Intellectual status and intellectual growth. Journal of Educational Psychology, 1966, 57, 121-127.

Whittlesey, J. R. B. A comparison of the correlational behavior of random number generators for the IBM 360. Communications of the Association for Computing Machinery, 1968, 11, 641-644.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Footnote

[1]The author wishes to acknowledge the assistance of Robert T. Patrick, who wrote the computer programs for analyzing the data.

Table 1

Variables Used in the Study

Variable No.	Description
1	K-12 District Instructional Expense per Pupil (1969-70)
2	K-12 District Current Operating Expense per Pupil (1969-70)
3	Percent of Non-White Students
4	Pupil/Professional Instructional Staff Ratio (includes teachers, principals, librarians, counselors, etc.)
5	Pupil/Teacher Ratio
6	Percent of Teachers with Five or More Years of Experience
7	Percent of Teachers with Master's Degree
8	Total Number of Students in School (Fourth Friday counts)
9	Total Number of Third-Graders (Fourth Friday counts)
10	State-Funded Compensatory Education Program? (Yes, No)
11	State-Funded Remedial Reading Program? (Yes, No)
12	Pretest Total Reading Mean
13	Pretest Total Reading Standard Deviation
14	Posttest Total Reading Mean
15	Posttest Total Reading Standard Deviation
16	Weekdays between Pretest and Posttest
17	Percent of Students Participating in Title I Reading
18	Percent of Students Taking Both Tests
19	Within-School Regression SEI (Low-Scoring Students)
20	Within-School Regression SEI (Middle-Scoring Students)
21	Within-School Regression SEI (High-Scoring Students)
22	Within-School Regression Standard Error of Estimate
23	Within-School Regression Corrected SEI (Low-Scoring Students)
24	Within-School Regression Corrected SEI (Middle-Scoring Students)
25	Within-School Regression Corrected SEI (High-Scoring Students)
26	Individual Residual SEI
27	Individual Residual Standard Deviation
28	Mean Difference Score
29	Mean Difference Standard Deviation
30	School Residual SEI (Pretest)

Table 2

Within-School Regression Coefficients and Other Information for Schools That Had the Highest and Lowest School Effectiveness Indices for High-, Middle-, and Low-Scoring Students[a]

	High-Scoring Students		Middle-Scoring Students		Low-Scoring Students	
Item	Highest SEI	Lowest SEI	Highest SEI	Lowest SEI	Highest SEI	Lowest SEI
School Code	082	152	082	152	192	042
No. of Students	28	39			46	143
High Pretest Score	61	65			94	81
Low Pretest Score	38	8			36	13
Pretest Mean	47.5	42.7			59.5	52.2
Pretest SD (N)	6.9	10.3	Same as for High-Scoring Students		11.9	9.6
Posttest Mean	58.8	46.5			67.8	55.2
Posttest SD (N)	11.1	9.1			9.5	10.3
Pretest-Posttest Correlation	0.90	0.78			0.59	0.86
Estimated Pretest Reliability	0.92	0.97			0.98	0.96
Intercept (uncorrected)	-9.65	16.92			39.99	6.93
Slope (uncorrected)	1.44	0.69			0.47	0.92
Intercept (corrected)	-15.19	15.89			39.26	4.96
Slope (corrected)	1.56	0.72			0.48	0.96
SEI (uncorrected)	80.8	60.5	65.0	52.9	59.1	44.7
SEI (corrected)	82.6	61.0	65.6	53.1	58.9	44.3

[a]The reference points were 62.7510, 51.8201, and 40.8892 for high-, middle-, and low-scoring students, respectively.

Table 3

Regression Coefficients and Other Information on the Individual Residual and School Residual Models

Item	Model: Individual Residuals	Model: School Residuals (Pretest)
Response Variable	Posttest Score	School Mean Post-test Score
Predictor Variable	Pretest Score	School Mean Pre-test Score
No. of Observations	3769	70
Predictor Mean	51.82	51.82
Predictor SD	10.93	9.87
Response Mean	58.34	58.54
Response SD	11.27	10.14
Predictor-Response Correlation	0.80	0.90
Intercept	15.60	8.98
Slope	0.82	0.96
Standard Error of Estimate	6.76	2.14

Table 4

Intercorrelations among School Effectiveness Indices Estimated by Five Different Methods

	19	20	21	23	24	25	26	28	30
19		0.8120	0.3166	0.9979	0.8043	0.2764	0.7853	0.6579	0.6997
20			0.8107	0.7924	0.9982	0.7833	0.9471	0.8245	0.8674
21				0.2869	0.8154	0.9957	0.7515	0.6800	0.7077
23					0.7863	0.2460	0.7677	0.6507	0.6898
24						0.7923	0.9462	0.8422	0.8805
25							0.7263	0.6787	0.7002
26								0.9301	0.9608
28									0.9955
Mean	49.77	58.64	67.51	49.41	58.66	67.92	0.20	6.73	0.00
S.D.	2.99	2.42	2.98	2.99	2.38	3.03	2.20	2.12	2.11

Table 5

Derived Minres Factors for School Effectiveness Indices

(Normal Varimax)

Variable	Factor I	Factor II	Factor III	h^2
1. W/Schl Reg. (L)	.95	.12	.28	1.000
2. W/Schl Reg. (M)	.67	.66	.35	1.000
3. W/Schl Reg. (H)	.13	.95	.28	1.000
4. W/Schl Reg. Corr. (L)	.95	.09	.28	.997
5. W/Schl Reg. Corr. (M)	.65	.66	.37	.998
6. W/Schl Reg. Corr. (H)	.08	.95	.30	.995
7. Ind. Residuals	.58	.53	.59	.969
8. Mean Differences	.40	.42	.81	.994
9. School Residuals	.45	.46	.77	1.000
Factor Variance	3.42	3.36	2.17	
% Factor Variance	38.2	37.5	24.2	
% Total Variance	38.0	37.3	24.1	
Eigenvalue	6.93	1.54	0.48	

Table 6

Correlations of the Various Types of School Effectiveness Indices with 21 Other Variables

	1	2	3	4	5	6	7	8	9	10
19	0.00	-0.01	-0.33	-0.13	-0.05	0.10	0.19	-0.20	-0.22	-0.08
20	0.09	0.09	-0.46	-0.16	-0.09	0.08	0.15	-0.31	-0.31	-0.08
21	0.14	0.16	-0.42	-0.12	-0.10	0.04	0.05	-0.30	-0.28	-0.04
23	-0.01	-0.01	-0.31	-0.14	-0.05	0.07	0.18	-0.20	-0.21	-0.07
24	0.10	0.11	-0.43	-0.18	-0.11	0.06	0.15	-0.31	-0.31	-0.04
25	0.16	0.18	-0.37	-0.15	-0.13	0.03	0.05	-0.29	-0.28	0.00
26	0.18	0.19	-0.37	-0.18	-0.09	-0.01	0.18	-0.25	-0.25	-0.06
28	0.25	0.32	-0.14	-0.31	-0.20	-0.15	0.20	-0.20	-0.22	0.11
30	0.23	0.29	-0.20	-0.28	-0.17	-0.12	0.19	-0.22	-0.23	0.07
Mean	541.73	789.11	33.83	22.19	25.53	57.56	19.57	492.84	68.26	0.23
S.D.	63.46	95.34	35.19	4.29	3.88	15.58	11.16	258.75	38.73	0.42

	11	12	13	14	15	16	17	18	22	27	29
19	0.11	0.41	0.22	0.67	0.03	-0 06	-0.09	0.26	0.65	0.67	0.74
20	0.13	0.41	0.14	0.74	0.33	-0.03	-0.08	0.30	0.57	0.62	0.56
21	0.10	0.26	0.00	0.54	0.51	0.01	-0.04	0.22	0.28	0.34	0.17
23	0.12	0.38	0.26	0.64	0.04	-0.05	-0.08	0.25	0.66	0.68	0.76
24	0.13	0.36	0.12	0.71	0.32	-0.02	-0.06	0.27	0.56	0.61	0.55
25	0.09	0.19	-0.06	0.48	0.46	0.02	-0.02	0.19	0.23	0.29	0.12
26	0.06	0.28	0.08	0.66	0.30	-0.01	-0.09	0.26	0.57	0.61	0.55
28	0.02	-0.10	-0.04	0.34	0.22	0.10	0.04	0.10	0.48	0.52	0.45
30	0.03	-0.00	-0.01	0.43	0.24	0.07	0.01	0.14	0.51	0.55	0.48
Mean	0.46	51.82	9.87	58.54	10.14	135.30	21.50	78.31	6.26	6.30	6.57
S.D.	0.50	4.63	1.74	4.91	1.79	9.09	26.18	11.55	1.38	1.45	1.61

Table 7

Variance Components, Signal-to-Noise Ratios, and Reliability Coefficients for School Effectiveness Indices Estimated by Five Different Methods

	Measure	Variance Component σ^2schools	σ^2samples	Signal/ Noise Ratio	Reliability Coefficient
19.	Within-School Regression SEI (Low-Scoring Students)	7.42	4.44	3.34	.77
20.	Within-School Regression SEI (Middle-Scoring Students)	4.80	2.45	3.92	.80
21.	Within-School Regression SEI (High-Scoring Students)	5.76	7.77	1.48	.60
23.	Within-School Regression Corrected SEI (Low-Scoring Students)	7.31	5.23	2.80	.74
24.	Within-School Regression Corrected SEI (Middle-Scoring Students)	4.50	2.54	3.54	.78
25.	Within-School Regression Corrected SEI (High-Scoring Students)	5.34	9.52	1.12	.53
26.	Individual Residual SEI	4.16	1.48	5.61	.85
28.	Mean Difference Score	3.81	1.53	4.98	.83
30.	School Residual (Pretest) SEI	3.76	1.50	5.01	.83

Table 8

Regression Coefficients for Predicting the Within-School Regression Corrected School Effectiveness Index (Middle-Scoring Students) from Various Predictor Subsets

Predictor	No. of Predictors (p) 1	2	3	4	5	6	7	8	9	Correlation of Predictor with SEI	Multiple Correlation with the Other Eight Predictors
14'	.34	.36	.38	.35	.37	.36	.35	.37	.37	.70	.38
[illegible]		-.16	-.15	-.16	-.16	-.17	-.15	-.14	-.15	-.18	.36
2			.01	.01	.01	.01	.01	.01	.01	.11	.76
15'				.22	.21	.22	.21	.20	.20	.34	.17
1					-.01	-.01	-.01	-.01	-.01	.10	.74
7						.02	.03	.03	.03	.15	.17
8							-.00	.00	.00	-.31	.87
18								-.02	-.02	.27	.40
9									.01	-.31	.87
Constant	39.02	41.39	35.48	35.30	35.41	36.11	36.67	37.46	37.43		
R	.70	.76	.79	.80	.81	.82	.82	.83	.83		
SE_{est}	1.72	1.58	1.50	1.47	1.45	1.44	1.43	1.44	1.44		
$^*F_{1,69-p}$	64.98	13.48	8.36	3.83	3.14	2.15	1.28	0.67	.33		
*Pr	<.001	<.001	<.01	<.10	<.10	>.10	>.10	>.10	>.10		

*F and Pr refer to the last predictor entered into the regression equation.

Figure Captions

Fig. 1. Within-school regression lines for three hypothetical schools.

Fig. 2. Comparison of "true" and observed regression lines.

Fig. 3. Plot of school effectiveness indices for middle-scoring students from corrected within-school regression and school effectiveness indices from total individual residuals.

Fig. 4. Plot of school effectiveness indices for middle-scoring students from corrected within-school regression and school effectiveness indices predicted from three variables.

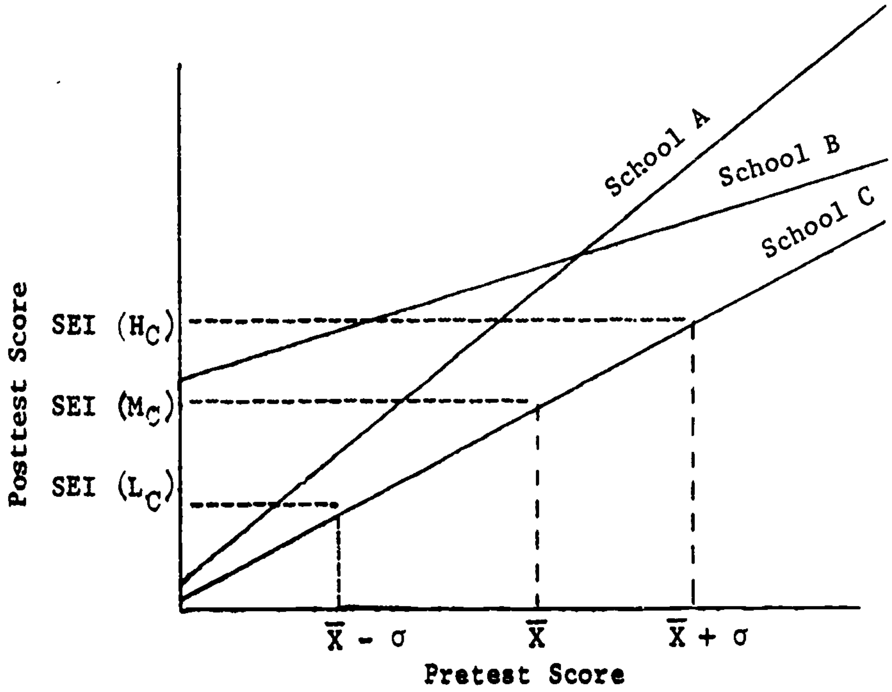

Fig. 1. Within-school regression lines for three hypothetical schools.

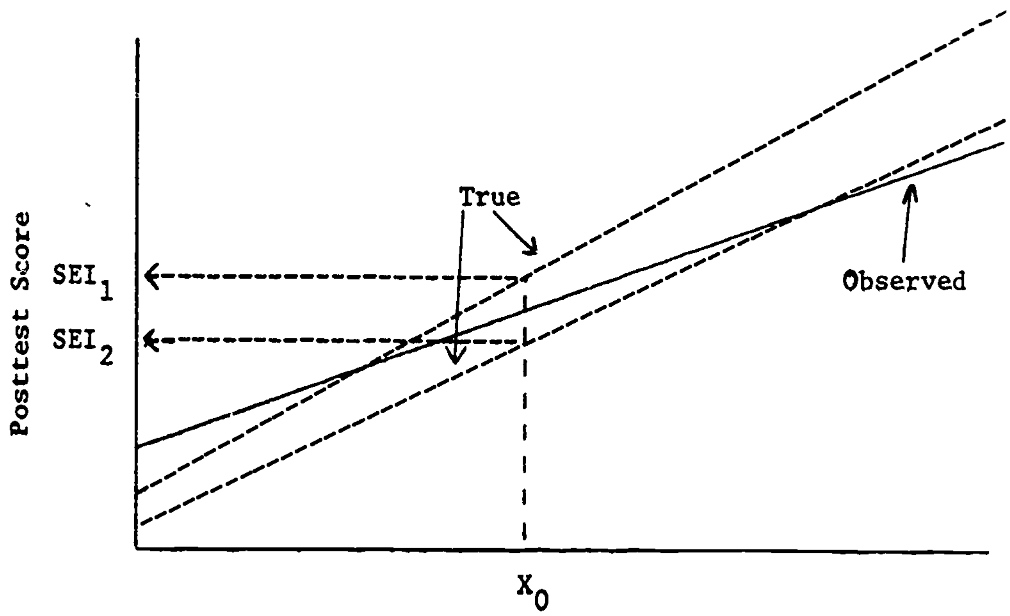

Fig. 2. Comparison of "true" and observed regression lines.

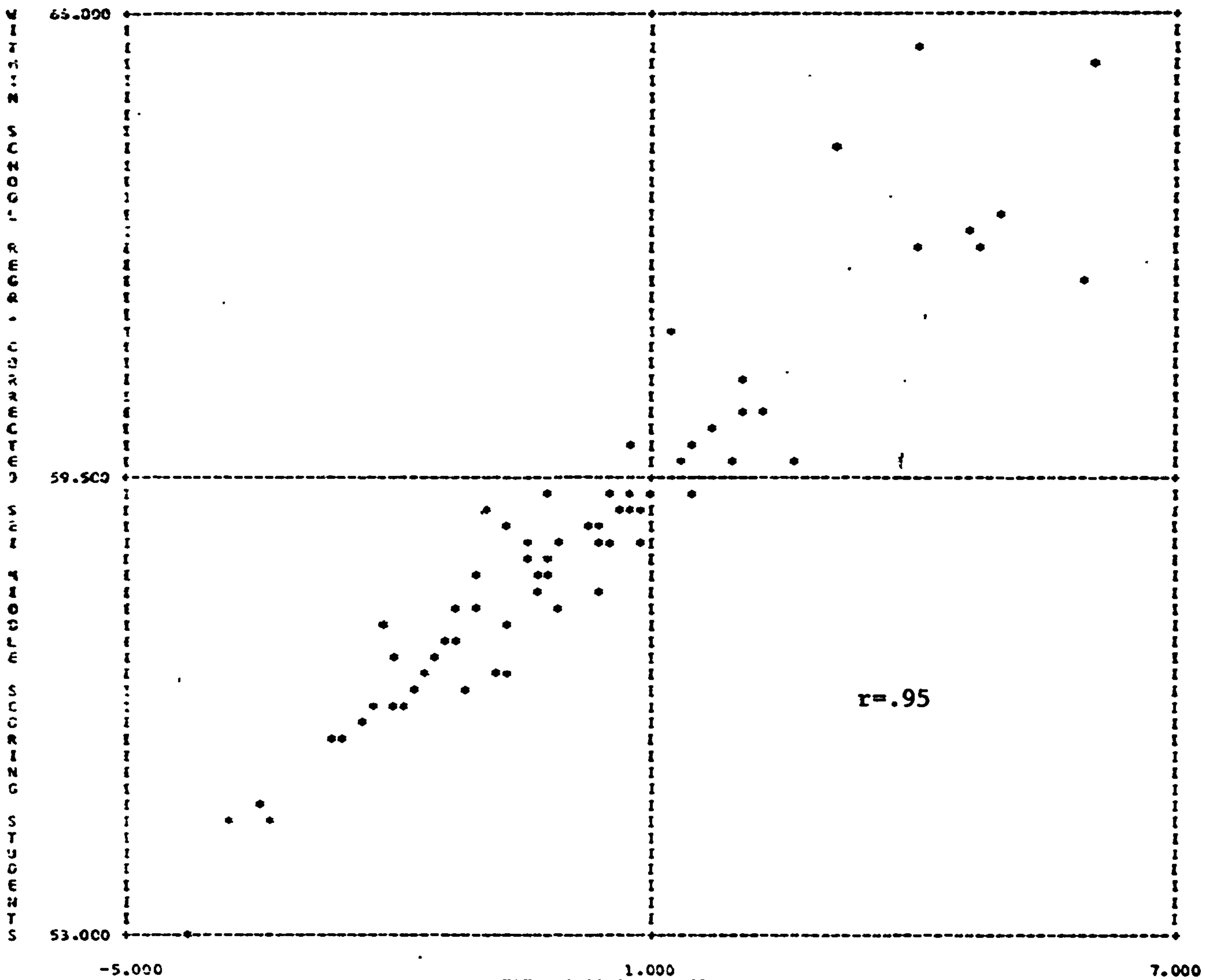

Fig. 3. Plot of SEIs for middle-scoring students from corrected within school regression and SEIs from total individual residuals.

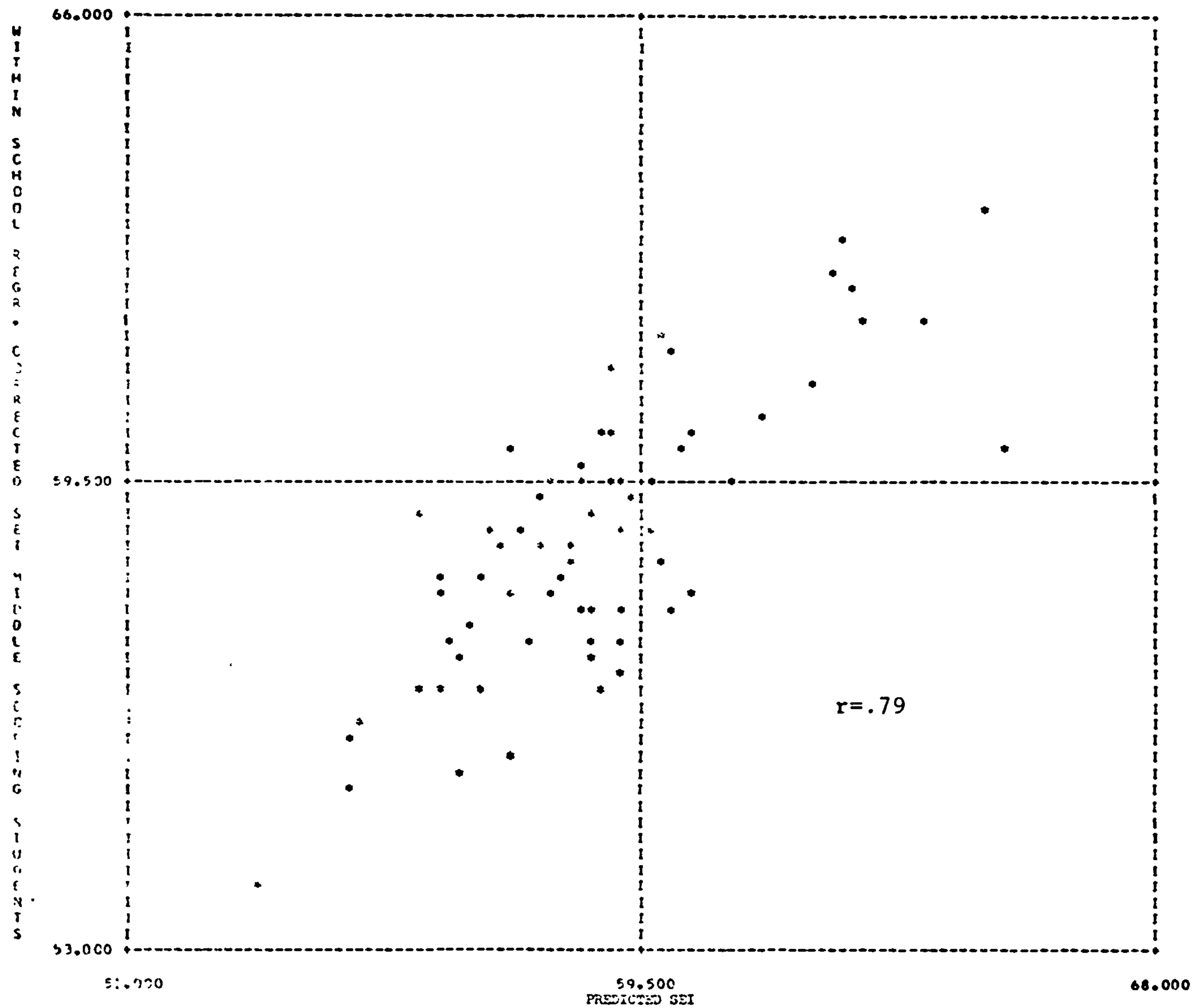

Fig. 4. Plot of SEIs for middle-scoring students from corrected within-school regression and SEIs predicted from three variables.