A COMPARISON OF THE STABILITY OF SCHOOL EFFECTIVENESS INDICES PRODUCED BY CLASSICAL LEAST SQUARES REGRESSION AND BAYESIAN M-GROUP REGRESSION TECHNIQUES

by

John R. Endahl

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTORATE OF PHILOSOPHY

in

Educational Research and Evaluation

APPROVED:

Dennis E. Hinkle

Lawrence H. Cross

Jim C. Fortune

Lee M. Wolfle

Wayne M. Worner

March, 1983
Blacksburg, Virginia

A COMPARISON OF THE STABILITY OF SCHOOL EFFECTIVENESS INDICES PRODUCED BY CLASSICAL LEAST SQUARES AND BAYESIAN M-GROUP REGRESSION TECHNIQUES

by

John Raymond Endahl

(ABSTRACT)

Numerous school effectiveness studies have utilized least squares regression techniques to produce school effectiveness indices despite the fact that they are subject to serious sampling fluctuations when sample sizes are small. If the sample size is smaller than normally thought adequate for accurate prediction a larger sample can be analyzed by pooling students from similar programs from different schools. Even though the regression weights for similar programs should be similar across schools, direct pooling of students may be less than satisfactory. A technique such as Bayesian m-group regression can be used that will incorporate both the similarity of the regressions across schools as well as the uniqueness of the individual programs.

This study empirically examines the predictive efficiency of four regression techniques that utilize individual student data as input. Cross-validation analyses were performed and mean squared errors, mean absolute errors, and correlations between observed and predicted scores were compared for four methods: (1) within-school least squares regression, (2) pooled least squares regression, (3) pooled least squares regression with adjusted alphas, and (4) Bayesian m-group regression with identical regression coefficients.

In addition, school effectiveness indices were obtained for the four regression techniques as well as least squares regression using school means and mean difference scores. These effectiveness indices were compared, and the stability of these indices across random samples of students, and across consecutive classes examined.

The within-school least squares regression method was found to be somewhat inferior to the other three models in terms of predictive efficiency. The Bayesian m-group equal slope model showed no appreciable advantage over the pooled least squares regression model or the pooled least squares regression model with adjusted alphas.

The indices produced by all six methods appear to be capable of representing the relative effectiveness of the schools involved in the study. In addition, those indices that moderate the importance of extreme values remained relatively stable from one subsample to another with correlations ranging from .75 to .85. Stability from class to class were of a much lower magnitude than those values reflecting stability from sample to sample. Correlations between school effectiveness indices of consecutive classes ranged from .28 to .47.

ACKNOWLEDGEMENTS

The author would like to take this time to thank the many people who never let me forget that there was indeed light at the end of the tunnel. I am forever grateful to my committee members whose contributions to my educational career extended far beyond the classroom. Special thanks to Dr. Dennis Hinkle for serving as the chairman of the committee and being so patient when the first attempt did not meet my expectations. Thanks to Dr. Jim Fortune whose friendship and intellectual stimulation has been on ongoing process. I also wish to thank Dr. Lee Wolfle whose organization is a constant reminder that a researcher's desk need not always be cluttered; Dr. Lawrence Cross whose humor in the classroom has shown me that research courses need not be dry and uninteresting; and to Dr. Wayne Worner who provided a much needed practical viewpoint to the study. To each of these committee members I take off my hat to you and thank you for bearing with me and serving on my committee until the end.

I would also like to thank for providing the data with which to test my research questions.

I am indebted to my wife, , who never lost faith in me and with her constant words of encouragement, love and devotion supported me in the long struggle to finish. Words cannot express how much she has meant to me in this endeavor.

Special thanks are extended to my loving parents, who have encouraged and supported me at every step in my educational progression.

Finally a note of thanks to those special people who have given me constant support throughout this long process, and who have made me aware that this is not the termination of an educational dream, but only just the beginning.

TABLE OF CONTENTS

LIST OF TABLES

Chapter I

INTRODUCTION AND STATEMENT OF THE PROBLEM

INTRODUCTION

The concept of professional accountability in educational settings has come under close scrutiny in recent years. Educators are being asked to provide information to the public indicating how well the school staff is fulfilling its professional obligations in working toward some clearly defined pupil performance objectives. Government officials seek accountability information as a prerequisite for new appropriations. To the layperson collecting information to be used as a basis for measuring the effectiveness of educational systems is a simple process involving (1) administering a commonly used standardized test of academic achievement, (2) reporting the average test score of students within a given system, and (3) comparing these average achievement test scores with a local or national average or norm. This approach assumes that the discrepancy between the two averages measures the educational effectiveness of the system. Under this strategy, those schools above the norm are performing effectively while those schools below the norm are, for some reason, falling short.

To the professional educator, however, there are a number of well-known fallacies in this type of discrepancy measure as an index of educational effectiveness. Using the discrepancy measure of system performance does not take into account the achievement level at which students enter into the system nor the conditions in which the schools must operate. It is entirely possible that a school reporting a lower than average test score at the end of an instructional period may have achieved more impressive changes than a second school with higher test averages if the characteristics of pupils in these two schools are markedly different. In order to make more rational comparisons of measures of school system effectiveness, it is important to consider differences in the characteristics of pupils in all schools prior to instruction.

Issues related to the measurement of change, including adjustments for differences in pupil characteristics prior to instruction, have been addressed by many researchers (e.g. Campbell & Erlebacher, 1970; Cronbach & Furby, 1970; Dyer, 1970; Dyer, Linn & Patton, 1967; Elashaff, 1969; Linn & Slinde, 1977; Lord, 1967; Porter, 1967; Werts & Linn, 1971). Although the debate continues as to the appropriateness of the various methods, practically all involve some adjustment of the final achievement measures based on measures taken prior to the instructional period.

One of the better known approaches to developing an accountability system is the Student Change Model proposed by Dyer (1967). Dyer proposed the use of multiple linear regression techniques for obtaining descrepancy measures based on observed and predicted school system achievement means. Predictor variables used in the regression analyses included prior achievement measures and measures of "hard-to-change" surrounding conditions (e.g. student's socioeconomic status, community wealth, and pupil-teacher ratio) and educational process variables (e.g. teaching methods, counseling).

A model similar to the Dyer model, the Production Process Model, has been used by economists to investigate school effectiveness (Hanushek, 1970; Winkler, 1975). In this model, educational outcomes are viewed as a function of initial student ability or achievement, family background measures, peer characteristics, community influences, and school influences. These functions are used to identify variables that can be manipulated which affect achievement rather than to identify effective schools. The Production Process Model thus includes more variables to determine outcomes, while the Dyer model uses only input and hard-to-change variables.

Although, ideally, the predictors of school effectiveness should include longitudinal achievement data and demographic measures, it is often impractical to obtain data necessary for the implementation of such a complete model. Recent studies of the relative effectiveness of schools (Burk, 1978; Gastright, 1977; Convey, 1974) have compared effectiveness measures based solely on achievement variables, solely on nonachievement variables, and a combination of the two types of variables.

Additional school effectiveness studies (Burk, 1978; Hilton and Patrick, 1970; Dyer, Linn and Patton, 1969) have compared the effectiveness indices based on matched-longitudinal achievement data, unmatched-longitudinal data, and/or cross-sectional data. Although the matched-longitudinal approach is the least practical, it appears that this approach is the most valid method for measuring school effectiveness.

Studies by Dyer, Linn and Patton (1969) and Forsyth (1973) indicate that a matched-longitudinal method, which measures performance based on the same students at two points in time, is most likely to provide valid measures of system effectiveness. Discrepancy measures of school system performance based on means computed from all cases present at pretest and posttest administration may lead to invalid conclusions regarding school comparisons. A major problem associated with measures that rely on longitudinal data based on the same students is the attrition in the student sample. If the mobility of the student population is relatively high, results may be based on relatively small samples. Therefore, the validity of school performance indices will depend largely on the efficiency of the procedures in utilizing small samples for estimating the individual adjusted output scores.

The multiple regression analyses used for a majority of the school effectiveness studies have used least squares techniques despite the fact that they are subject to serious instability when sample sizes are small. If the sample size is smaller than normally thought adequate for accurate predictions a larger sample can be analyzed by pooling students from similiar programs from different schools. Although the regression weights for similiar programs should be similiar across schools, direct pooling of students will be less than satisfactory. A technique is needed that incorporates both the similarity of the regressions across schools and also the uniqueness of the individual school programs.

When the same predictions must be made for several groups, (i.e., schools within a school district) improvements in prediction are possible. Jackson, Novick and Thayer (1971) presented a Bayesian method, based on theory by Lindley (1970), for the simultaneous estimation of regression lines in m groups, which _a priori_ are believed to have similar characteristics. By assuming the property that de Finetti (1964) has called "exchangeability," Bayesian estimates of the regression parameters utilize collateral information available in the other (m-1) groups. This Bayesian m-group regression approach has been shown to be particularly effective in predicting the level of performance for individuals in a number of institutions based on standardized tests (Jackson, Novick, and Thayer, 1971; Novick, Jackson, Thayer, and Cole, 1972; Novick and Jackson, 1974; Shigemasu, 1976).

The great strength of the Bayesian model is that many classical models are encompassed in the Bayesian model as special cases and thus do not initially commit the researcher to any one particular model. Depending on the evidence from initial data analyses, the Bayesian model selected incorporates a mixture of these models to provide a regression line that converges to the true regression line.

Shigemasu (1976) extended this theory by introducing the Bayesian m-group regression model with identical regression coefficients. This model assumes that "the variances across groups of the individual regression coefficients, except for intercepts, are zero, i.e., all regression coefficients across groups are presumed identical (Shigemasu, 1976, p. 158)." Results from analyses on two extensive data bases indicated that the Bayesian m-group regression analysis with identical regression coefficients provided almost as good a prediction as the full m-group regression analysis. Considering the substantial reduction in computational costs, this model showed promise.

However, not all studies which compared Bayesian and non-Bayesian regression estimates have indicated that Bayesian m-group regression analysis offered improvement over the classical least squares method (Lissitz and Schoenfeldt, 1974; Houston, 1976; Polhamus, 1980). Shigemasu (1976) suggested that a number of diverse application studies be done to examine which situations fit the model.

Novick et al. (1972) indicated that, in theory, the Bayesian simultaneous estimates should provide more accurate prediction functions on the average than those using the least squares estimates. To date, relatively few cross-validation studies have been undertaken using real data as opposed to computer-generated data to examine the predictive efficiency of the models. The need for additional cross-validation studies in order to check shrinkage of the multiple correlations has been stressed by Novick and Jackson

(1974) and Kerlinger and Pedhauzer (1973), among others. Only a cross-validation study "with real data can pinpoint precisely how a specific technique will work on further data sets" (Novick et al., 1971, p. 4).

STATEMENT OF THE PROBLEM

The debate as to the usefulness of Bayesian m-group regression in an educational setting continues. Additional studies need to be undertaken before this question is resolved. Coffman and Shigemasu (1978) have suggested that Bayesian m-group regression techniques might be utilized in school effectiveness studies. If residual scores produced by either Bayesian or non-Bayesian techniques are to be used as measures of school effectiveness it is imperative that the prediction models be as precise as possible. Therefore, the need exists for additional studies which examine the predictive efficiency of both classical regression models and Bayesian m-group regression models.

Regardless of the method used to produce school effectiveness indices, whether it be a classical approach or a Bayesian approach, the reliability of these effectiveness indices must be examined before these indices are used to identify exceptional schools. Differences in school effectiveness measures should represent more than just error variation.

One type of stability that might be considered important is the consistency of the effectiveness measures across random samples of students from the same grade level of the local educational system. Dyer, et al. (1967) proposed that within each school the student population be randomly subdivided into halves, and multiple regression procedures be applied independently to the two subsamples. Correlations of the residuals using classical regression techniques for these two independent samples were found by Dyer, et al. (1969) to have a median correlation across subtests of .70. This type of reliability estimate is concerned with pupils as a source of error.

Marco (1974) compared the stability of a number of school effectiveness indices including mean difference scores, individual residual scores, and school residuals across samples using two random halves for each of 70 elementary schools. Reliability coefficients for the school effectiveness indices were computed by means of analysis of variance. All but one of the school effectiveness indices appeared to be stable enough to be used as a measure of school effectiveness. Individual residuals were the most stable across samples with a reliability coefficient of .85.

A second type of stability index examines the consistency of indices for consecutive classes in the same school. Forsyth (1973) predicted mean school twelfth grade achievement scores using mean school ninth grade achievement scores for unmatched longitudinal samples. Forsyth reported multiple correlation coefficients similar to those reported by Dyer et al. (1969). However, the correlations between residuals for the consecutive years (.11 to .50 with a median r=.28) were considerably lower than the random halves correlations reported by Dyer. Acland (1972) and Gastright (1974) reported somewhat higher correlations among residuals for consecutive classes using unmatched student groups. In addition, Gastright (1977) provided no empirical evidence that the addition of nonachievement variables increased the stability of residuals across years.

RESEARCH QUESTIONS

The use of standardized tests of academic achievement as a basis for measuring the effectiveness of educational systems has had a long history and will continue to be used in school effectiveness studies. The models that use this information in the development of school effectiveness measures differ considerably. The purpose of this study is to investigate empirically a number of questions that arise out of the development of reliable measures of school effectiveness. Initial questions examine the accuracy of the predictions of regression models using individual student information as predictors. A second set of questions examine the comparability of school effectiveness indices produced by regression models that use individual student information as well as two additional models. The last two sets of questions examine the stability of these effectiveness indices across samples and across consecutive classes. The research questions for this study are as follows:

Question 1. To what extent are classical and Bayesian regression models comparable when applied to individual, matched-longitudinal samples? How comparable is the predictive efficiency of these regression models. The four regression models used in this comparative study are as follows:

Method 1: Within-school least squares regression.

Method 2: Pooled least squares regression.

Method 3: Pooled least squares regression with adjusted alphas.

Method 4: Bayesian m-group regression with identical regression coefficients.

Question 2. To what extent are school effectiveness indices comparable when obtained by each of six different methods? The six methods that are used in this comparative analyses include those methods listed in Question 1 as well as the following:

Method 5: Least squares regression using school means scores.

Method 6: Mean difference scores.

Question 3. How stable are these measures of school effectiveness across random samples of students from the same grade level of the schools.

Question 4. How stable are these measures of school performance for consecutive classes assuming no "out-of-the-ordinary" changes in the educational programs have been made?

Chapter II

REVIEW OF THE LITERATURE

CLASSICAL MODELS OF ACCOUNTABILITY

Perhaps the best known classical regression approach to developing accountability systems was one proposed by Dyer (1970). Prior to popularization of the term accountability, Dyer suggested a rationale for measuring school system effectiveness based on residual mean performance for schools. Within his Student Change Model of a school, Dyer (1970) defines four groups of variables existing in the educational system which must be measured in order to develop some criteria for accountability of the school skills. These four major categories of variables are called input, educational process, surrounding conditions, and output. The input and output variables consist of characteristics of the pupils measured before and after a given period of instruction. These variables may include such characteristics as personal health, academic achievement, attitudes toward school, educational aspirations and education motivation. Implementation of these input and output variables usually consists of pretest scores as input measures and posttest scores as output measures. In this context, the pupil input represents a fixed condition which the school staff has no control over. However, the pupil output reflects to some extent the quality of education provided by the school.

The third group of variables in the Student Change Model consists of surrounding conditions which represent factors in the school environment which may potentially have some effect, either positive or negative, on the education of pupils. These variables fall into three groups: home conditions such as parent's education, the level of family income, and parental pressures; community conditions including population density, ethnic composition of the enrollment area, and number of social agencies available; and school conditions which includes such things as pupil-teacher ratio, quality of the school plans, classroom footage per pupil, and the like. Surrounding conditions for which the school staff has little or no control over are labeled "hard-to-change" surrounding conditions while those conditions which the staff potentially has the capabilities to change are labeled "easy-to-change" surrounding conditions.

Finally, the educational process variables consist of all activities of the school expressly designed to stimulate student achievement. These variables include such activities as instructional processes, parental consultation, and vocational counseling. This group of variables combined with the other three groups of variables previously described, interact with one another in a complex manner. A composite index that synthesizes this complex relationship between input variables, educational process variables, surrounding variables, and output variables would be invaluable to school staff. By examining a readily interpretable composite index, school staff can know how well it is doing in its efforts to produce desired changes in pupils.

These indices are what Dyer (1970) refers to as "School Effectiveness Indices." These indices are computed by obtaining school means for each measure of input, output and hard-to-change surrounding conditions. Given output scores are regressed on the input measures and the hard-to-change surrounding conditions to produce predicted output scores. The discrepancies between the predicted output scores and the actual output scores are used as measures of school effectiveness. Schools with posttest scores above the predicted posttest scores would receive favorable school effectiveness indices, while schools with observed posttest scores lower than predicted would receive relatively low indices of school effectiveness.

Once these school effectiveness indices are obtained, the easy-to-change surrounding conditions and the educational process variables are examined for the outliers, i.e., those schools with extremely high or low school effectiveness indices. The comparison of these outliers for a given pretest condition would provide some insight into the identification of variables which might produce better student performances.

Dyer, Linn and Patton (1970) provided results related to the methods of obtaining discrepancy measures used as indices of school effectiveness. They compared four methods of obtaining discrepancy measures: (1) regressions of individual student output on individual student input for a matched-longitudinal sample, (2) regression of mean school system output on mean school system input for the same matched-longitudinal sample; (3) regression of mean school system output on mean school input for an unmatched-longitudinal sample; and (4) regression of mean school system output on mean school system input for a cross-sectional sample. Discrepancy measures of school system effectiveness obtained by methods 1 and 2 were found to be essentially interchangeable but those measures obtained by methods 3 or 4 were not considered suitable substitutes for effective measures obtained using the first two methods.

Dyer, Linn and Patton (1970) also examined the stability of the discrepancy measures produced by method 1 (individual, matched-longitudinal sample) and method 2 (system means, matched-longitudinal sample). Students were randomly subdivided into two halves and residuals were computed for each subsample. The correlations based on individual student scores for each subsample ranged from .73 to .88 for six subtests with a median of .78. The correlation between residuals of the two subsamples based on system means ranged from .62 to .84 for the six subtests with a median of .72. The residuals obtained from both methods remain relatively stable from one subsample to another subsample within a given year, with the residuals based on individual scores being slight more stable.

Forsyth (1973) examined the stability across time of school effectiveness indices obtained by Dyer's model. Forsyth used two matched-longitudinal samples from two time periods: 1965-68 and 1966-69. Using school means of the Iowa Test of Educational Development (ITED) as input and output variables, nine subtest scores and a composite score for 1965 were used as predictors of each at the ten ITED means for 1968. Similar predicted means for 1969 were obtained using 1966 ITED means as input. Forsyth (1973) found school residuals to be relatively unstable from one year to the next. The correlations between the residuals from those two prediction equations ranged from .11 to .50 for the nine subtests and one composite posttest, with a median correlation of .28.

Acland (1972) correlated residuals for consecutive classes using unmatched student groups. As the time between testing periods increased from one to three years, the intercorrelations between residuals for consecutive classes increased. These correlations between the residuals were interpreted by Acland as "a measure of the percentage of the variation in the residual scores that can be attributed to the stable characteristics of schools that raise performance levels."

Gastright (1974) examined the intercorrelations among residuals for three consecutive years using unmatched longitudinal student samples. School means on sixth grade reading achievement were predicted using mean school second grade reading achievement alone and coupled with community background variables collected concurrently with the second grade achievement testing. The results indicated that the residuals based solely on achievement input measures were more stable across the three year period than those obtained with the models that included nonachievement variables.

Marco (1974) compared five methods for computing school effectiveness indices using longitudinal data. Total Reading scores from the Metropolitan Primary Achievement Test were obtained as pretest and posttest measures of reading ability for third grades in 70 elementary schools. The five methods used to compute school effectiveness indices were: (1) within-school regression; (2) corrected within-school regression; (3) mean difference scores; (4) individual residual scores; and (5) school residual scores. All five indices yielded measures that were highly intercorrelated. However, the correlational patterns were somewhat different with other variables, and thus were not deemed interchangeable. With respect to the stability of these indices from sample to sample, most of the indices were highly stable with the individual residual score method yielding the most stable measures. Marco did not address the issue of stability over time.

Convey (1975) used simulated data for 54 hypothetical schools in an attempt to examine the effectiveness of three statistical models in reproducing indices that would reflect school effectiveness ratings established *a priori*. The three models used to produce indices were: (1) the within-school regression model where a prediction equation for each school was obtained by regressing individual student output scores on individual student input and SES scores, (2) the individual regression residuals model where one predictor equation was obtained for the entire group by regressing individual student output scores on individual input scores, SES scores, and a school variable which remained constant for all individuals in a given school, and (3) the school regression residuals model where one prediction equation was obtained for the total group using school means of the input scores, SES variables, and the school variable as predictor variables to predict mean school output. Each model was found to be capable of producing school effectiveness indices that rather accurately reflected the effectiveness ranks and effectiveness classifications established a priori. He concluded that in a cost-effectiveness sense, residuals based on school unit data should be considered superior. In addition, Convey concluded that in order to determine the relative effectiveness of schools, nonachievement variables should be included in the model specifications based either through theory, results of previous research, or personal insight.

Fillos and Bowman (1976) pointed out several weaknesses of Dyer's model. First the number of school districts having a sufficient number of schools to serve as a basis for a meaningful regression equation is relatively small. Second, by working with averages of all variables for a school a considerable amount of information may be lost or distorted. Finally, a statistical test of real differences between school performances using Dyer's model is not readily available. In an attempt to remedy these weaknesses, Fillos and

Bowman made the following adaptions of the Dyer model: "(1) to use regression as an exploratory rather than predictive technique, (2) to use individual student data rather than school averages, and (3) to include school membership as a nominal variable dummy coded in the regression analysis (p.3)." Variables in the study included grade equivalent scores from the Iowa Test of Basic Skills at both the fifth and seventh grades as well as background information such as sex, aptitude (Cognitive Abilities Test), self-reported occupation and education of both parents, and the school which the student attended. Stepwise regression analysis indicated that the contributions to explained achievement variance by socioeconomic status and sex were included in the information from previous achievement and aptitude measures.

Other studies have examined the comparability of school unit residuals based on achievement and nonachievement variables. Gastright (1977) correlated residuals based on matched-longitudinal data and unmatched-longitudinal data using only previous achievement data, achievement data and nonachievement data, and only nonachievement data as input variables. Residuals based on both sets of longitudinal data were highly correlated. However, residuals based on nonachievement variables were unrelated to those based on achievement alone. Gastright concluded that "the decision to use nonachievement input variables in the production of residuals as measures of school effectiveness cannot be made on the basis of empirical studies."

Burk (1978) investigated three commonly used school effectiveness models based on incomplete sets of data: (1) demographic predictors alone, (2) matched-longitudinal achievement predictors alone, and (3) unmatched-longitudinal achievement. Classifications using these three incomplete data sets were compared with classifications produced using the complete data set -- matched-longitudinal achievement predictors and demographic predictors. The correlation of residuals from the complete and incomplete models was greater than .70 with some similarity in the classification of schools with extreme residuals. Burk (1978) concluded that "incomplete models may be useful in school effectiveness studies conducted for the purpose of identifying only a small proportion of school as effective or ineffective, when failure to identify a school is not considered as serious as the incorrect identification of schools (p.89)."

Relatively few studies have utilized both demographic and previous achievement measures as predictor variables. The State of New York used unmatched longitudinal achievement and demographic variables to predict district mean test scores (N.Y. State Education Department, 1973). In 1975, the State of Maryland used two predictor models -- one using only ability as a predictor, the other using both ability and demographic predictors. The demographic variables were eliminated as predictors in the second year of implementation because the slight increase in the amount of achievement variance accounted for by the demographic predictors did not warrent the additional expense of collecting and analyzing this data.

BAYESIAN M-GROUP REGRESSION

In an attempt to improve predictions for small sample groups, Lindley (1970) extended the logic of Kelley's (1927) classical mental test theory of regression estimates of true score given the observed score. Kelley theorized that an estimate of a person's true score could be calculated by weighting the observed score for the person by the reliability of the test and adding this product to the product of the mean value of the test for all persons weighted by one minus the reliability of the test. Symbolically this can be written as follows:

$$\text{Estimated true score} = rx + (1-r)\bar{x}$$

where x is the person's observed score, $\bar{x}$ is the mean of all observed scores, and r is the reliability of the test.

If the reliability of the test defined as the ratio of the variance of true scores to the variance of the observed scores is relatively low, the collateral information obtained from the observed scores of others in the same school will have a profound influence on the estimation of a person's estimated true score. Kelly showed that the standard error obtained using the estimate above was substantially lower than the estimate produced by using only the observed score.

Lindley (1970) extended this notion with respect to using a Bayesian method for estimating regressions in m groups. Essentially, Lindley's theory determined simultaneous Bayesian regression equations which utilized collateral information available from the other (m-1) similiar groups to improve the prediction equation for each group.

More precisely, Bayesian m-group regression represents a weighted average of a particular group's classical regression line and the regression line obtained for all groups collectively. For each group of the m groups, classical least squares regression weights are computed for each of the p predictor variables. Similarly, classical regression weights are computed for each of the predictor variables for the groups pooled together. The Bayesian regression weights are then computed as a weighted average of the individual group regression weights and the pooled group regression weights.

For example, suppose the least square regression weights for group one were .08 for the first predictor variable and .03 for the second predictor variable. Similarly, the regression weights for the pooled sample of schools were .10 and .07 for the first and second predictor variables respectively. The Bayesian regression weights for group one would be:

$$(.08)(w1)+(.10)(1-w1) \text{ for predictor 1}$$

$$(.03)(w2)+(.07)(1-w2) \text{ for predictor 2}$$

The weights (w1,w2) used in the Bayesian m-group regression lines depend on the relative amount of information available for group one. If there exists a substantial amount of reliable information for this group then the weights (w1,w2) will be relatively large and the Bayesian m-group regression estimates will closely resemble the classical least squares regression estimates. If, however, group one has little reliable data, then the weights will be small and the Bayesian regression weights will resemble those produced by the regression of the pooled group.

Thus, the great strength of the Bayesian model is that it does not restrict the user to any one <u>a priori</u> model which may possibly be false. Depending on the weights attached to the direct information and the collateral information, the Bayesian model can resemble many classical models such as within-group least squares, pooled least squares and pool least squares with adjusted alphas. As the information for a particular group increases, the Bayesian estimated regression line will converge to the true regression line regardless of how this differs from the other true regression lines (Novick and Jackson, 1974, p. 79).

Applications of Bayesian Methods in Educational Prediction

The application of Bayesian methods in the educational setting began to generate interest with a study by Jackson, Novick and Thayer (1971) in which the feasibility of a Bayesian method for estimating regressions in m-groups was studied in an educational guidance setting. Bayesian m-group regression analyses were applied to data from Law School Admissions Test and a Comparative Guidance Program. The Bayesian slope estimates for most schools were quite different from those produced by the within-group least squares method. This diversity was "simply a reflection of the great inaccuracy with which regression coefficients are typically estimated and the substantial additional information provided by the Bayesian analysis (Jackson, Novick and Thayer, 1971, p. 142)." In each application, the Bayesian method was shown to produce plausible results.

Novick, Jackson, Thayer and Cole (1972) extended Lindley's Bayesian theory to multiple regression analysis and conducted a cross-validation study of the Bayesian m-group regression method. American College Test (ACT) subtest scores were used to predict first semester grade point average for students enrolled in 22 academic junior colleges. Classical within-group least squares regression estimates calculated from 1968 data were applied to 1969 data. Prediction accuracy of the two methods was examined using mean squared error. The cross-validation using the 100 percent 1968 sample yielded only modest predictive efficiency using the predictive efficiency using the Bayesian method. However, when prediction equations developed from a 25 percent sample were implemented on the 1969 sample, the Bayesian prediction equations produced a reduction of about 9.7 percent in mean squared error as compared with the within-group least squares equations. The Bayesian procedures tended to smooth out year-to-year sample variation. Comparing the mean squared error of prediction for a 25 percent sample using the Bayesian method with a 100 percent sample using within-group least squares, the results were essentially the same. The most apparent benefit of the Bayesian method is that it permits working with subpopulations with small sample sizes and where the regressions are different in the subpopulations.

Lissitz and Schoenfeldt (1974) conducted another cross-validation study using Bayesian and classical regression techniques in the context of the study of moderator variables. Unlike Novick et al.'s study (1972) which used natural pre-existing groups in the analysis, Lissitz and Schoenfeldt (1974) used groupings based on information obtained from a biographical inventory. The results of the crossvalidation which used high school grade point average, Scholastic Aptitude Test -- Verbal and Scholastic Aptitude Test -- Math to predict college grade point average were similar to those reported by Novick et al. (1972). Lissitz and Schoenfeldt concluded that the Bayesian procedure yielded results similar to those produced by the two probability weighting (semi-Bayesian) procedures and that the 10 percent reduction in mean square error reported by Novick et al. (1972) was due to "the comparison of the Bayesian procedure to a false standard, i.e., the within-group least squares results (p. 74)," since the within-group least squares model was consistently less effective than any of the models used in the study.

In response to Lissitz and Schoenfeldt's criticisms, Novick and Jackson (1974) compared the efficiency of the Bayesian predictions against two additional classical least squares models -- the pooled least squares model and the pooled least squares regression with adjusted alphas. In this study the Bayesian m-group regression method was found to be "meaningfully better" than the pooled least squares method and essentially equivalent to the pooled least squares regression with adjusted alphas.

One other study compared the Bayesian and least squares methods of educational prediction. Bolt (1975) used high school record, undergraduate record, and College Entrance Examination Verbal and Math subtest scores to predict grade point average. Very little variation in the predictive efficiency measures were produced using the two estimation methods.

Equal Slope M-Group Regression

In Novick and Jackson's (1974) summary of the status of Bayesian m-group regression, they noted that the simpler pooled least squares regression with adjusted alphas produced results essentially equivalent to the Bayesian method. Although the Bayesian approach was more difficult to operate, it did protect against "the inaccuracy of the model with respect to one or more of the groups." In an attempt to reduce the computational costs and time, Novick and Jackson (1974) suggested that "it is possible to do Bayesian regression in m-groups with the slopes assumed identical but with the intercepts subject to Bayesian estimation (p.83)."

Shigemasu (1976) developed this simplified m-group regression model and compared the predictive efficiency of this model with that of the full m-group regression and several least squares methods. Three data sources were used in the cross-validation studies including a 25 percent and 10 percent sample of 22 colleges' data found in the Novick et al. (1972) study. The full m-group regression model provided a slightly better prediction system than the Bayesian m-group regression analysis with identical slopes. Both Bayesian methods produced results similar to the pooled least squares model with adjusted alphas which were clearly superior to those produced by the within-group least squares method. "The Bayesian m-group regression with identical B also provided almost as good a prediction as the full m-group regression analysis in two extensive data bases. Considering the cost of computation, this method will be very useful for practical use (Shigemasu, 1976, p. 179)."

Two studies have used Shigemasu's equal slope m-group regression techniques (Houston, 1976; and Hinkle and Polhamus, 1980). In both studies grade point average was predicted using particular curriculum groupings offered in comprehensive community colleges as their m groups. Neither study found significant differences in the predictive efficiency of the Bayesian and classical regression model. It is quite possible that the equal slope model is not realistic when attempting to make predictions across diverse community college curricula.

Appraising School Effectiveness Using a Bayesian Method

Only one study has investigated the use of a Bayesian method to produce a reliable index of school effectiveness. Coffman and Shigemasu (1978) compared two methods of estimating adjusted mean scores, the classical analysis of covariance method and a Bayesian method that assumes exchangeability. Two successive cohorts of pupils from 19 elementary schools served as the samples for this cross-validation study. Scores from five subtests and the composite score of the Iowa Tests of Basic Skills were used as covariates to predict posttest scores taken two years later. The Bayesian estimates provided intuitively pleasing results and differed less from cohort to cohort than the classical estimates. Using Bayesian methods, only one school exhibited a positive deviation for one cohort and a negative deviation for the next cohort. Five such instances arose using classical estimates, with four of these cases involving vocabulary scores which are generally considered most resistent to change as a result of instruction.

Chapter III

RESEARCH METHODOLOGY

THE SAMPLE

Each year a large number of elementary schools in a suburban county of an eastern metropolitan area participate in a statewide testing program by administering the Iowa Test of Basic Skills (ITBS). A total of 120 elementary schools had administered these tests in 1976 and 1977 to their third grade students and also in 1978 and 1979 to their fifth grade students. A random sample of 50 schools were selected from those elementary schools.

Matched-longitudinal samples were formed for two time periods: 1976-78 and 1977-79. Cohort I (76-78) consisted of all students with identical student identification numbers having complete sets of 1976 third-grade test scores and 1978 fifth-grade test scores. At least 10 students from a single school had to have matched data before any of its students entered into the analysis. A total of 1946 students were included in Cohort I.

Similarly, Cohort II (77-79) was comprised of students with identical student identification numbers having complete sets of 1977 third-grade test scores and 1979 fifth-grade test scores. A total of 1933 students were included in Cohort II. The distribution of the number of students included in each cohort by school is found in Table 1.

THE VARIABLES

The variables selected for this study were similar to those used in previous school effectiveness studies. These variables included achievement variables as well as school and community background variables. Test scores from the Iowa Tests of Basic Skills were available for each student at both the third and fifth grade. Five major subtest scores and the composite score from the ITBS were used as achievement variables in the analysis. The five major subtests of the ITBS were: Vocabulary (V), Reading (R), Language (L), Work Study Skills (W), and Arithmetic (A). All test scores used in the analysis were Normal Curve Equivalent (NCE) scores. These scores are normalized standard scores with a mean of 50 and a standard deviation of 21.06. The NCE metric is an equal-interval scale with the length of the interval between any two adjacent scores on the scale equal to the interval between every other adjacent pair of scores. Since the NCE metric is an equal-interval scale it is legitimate to add, subtract, multiply, and divide NCEs. Therefore, a composite score was computed by adding the scores of the five major subtests together.

Table 1

Sample Sizes of Cohorts by School

School	Cohort I	Cohort II
51	63	60
101	24	23
102	47	33
106	35	55
205	26	31
206	52	39
207	25	38
209	44	40
220	24	44
225	12	23
226	72	56
235	74	57
302	30	35
304	23	15
307	11	16
308	53	49
310	59	66
312	38	42
313	63	37
409	13	13
414	32	38
415	24	28
416	19	12
422	32	53
504	46	37
505	63	59
509	41	18
552	45	43
553	38	37
559	32	33
561	37	58
563	37	45
566	25	15
570	53	62
601	65	68
604	41	36
652	26	27
767	33	35
768	18	22
769	53	38
772	43	52
774	44	45
785	32	38
786	25	25
788	65	47
795	35	30
797	35	50
808	35	28
821	52	45
822	32	37
TOTAL	1946	1933

For this study, the five major subtest NCE scores of the ITBS obtained during the third grade testing period were used as input variables in the regression analysis. The composite NCE score of the ITBS obtained during the fifth grade testing period was used as an output measure in the regression analysis.

In addition to the academic variables, numerous other variables were examined as variables that potentially could influence the quality of education. These background variables included school variables over which the school system had some policy control as well as community variables that were not amenable to policy control and could not readily be altered by school authorities. These variables were obtained from school records and appropriate school information was tagged to each individual student. A list of these background variables and their operational definitions can be found in Table 2.

THE MODELS

In order to examine the effectiveness and stability of a Bayesian method for estimating regressions in m groups compared with classical least squares regression methods, a number of regression models were used throughout the study. Four of these regression models used individual student test

Table 2

Operational Definitions of School and Community Variables

1. Average Pupil Attendance: The ratio of the average daily attendance to the average number belonging

2. Mobility Rate: The ratio of the total number of enteries and withdrawals during the year (78-79) to the official September 29, 1978 enrollment

3. Pupils Transported: The percent of pupils bused to school

4. Average Class Size: The ratio of total regular student enrollment to the number of attendance sections for this same group of pupils

5. Student/Staff Ratio: The ratio of total student enrollment to the total number of professional staff, excluding special education, Head Start, resource room or special needs teachers

6. 1979 Total Student Enrollment: The number of students enrolled on each grade, Head Start, and special education as of September 28, 1979

7. Total Minority: The percent of students enrolled on September 29, 1978 classified as American Indian, Asian, Black, or Hispanic

8. Staff Experience: The percentage of the staff members with less than 5 years of credited experience as recorded on the Personnel Master File

9. Staff Minority: The percentage of the staff members classified as American Indian, Asian, Black, or Hispanic as recorded on the Personnel Master File

10. Staff Sex: The percentage of male staff members as recorded on the Personnel Master File

11. Property Destruction: Total gross loss (including glass breakage, theft, miscellaneous destruction, and fire costs) for FY79

12. Special Education Availability: The percentage of students enrolled in self-contained special education classes.

scores as input and output variables. A fifth regression model used school means in the regression equations. A final model, school mean difference scores, served as a comparative base for the study. The following models were used throughout the study:

1. Within-school least squares regression. For each of 50 schools a regression line was computed that described the relationship between individual student pretest subtest scores and posttest composite scores.
2. Pooled least squares regression. Estimates of regression parameters for a single regression line were computed using individual student pretest subtest scores and posttest composite scores for the entire data set.
3. Pooled least squares regression with adjusted alphas. Regression coefficient estimates were calculated from individual student pretest subtest scores and posttest composite scores for the entire data set. The regression coefficients were assumed to be equal across schools but intercepts were estimated separately for each school.
4. Bayesian m-group regression with identical regression coefficients. Regression coefficient estimates were calculated by weighted averages using the direct in-

formation of individuals within a single school for one component of the average, and the collateral information contained in the mean value for all schools for the other component of the average. The regression coefficients were assumed to be equal across all schools but intercepts were estimated seperately for each school.

5. Least squares regression using school mean scores. For the entire group, a regression line was computed from the regression of the mean posttest composite scores for each school on the mean school pretest subtest scores.
6. Mean difference scores. For each school the mean difference score, ie., the posttest composite score minus the pretest composite score was computed.

SCHOOL EFFECTIVENESS INDICES

For each of the models used throughout the study at least one index was computed to examine school effectiveness. While the school effectiveness indices derived from the six models are not comparable in an absolute sense, the relative positions that the schools obtain on the various school effectiveness indices are comparable. Most of the indices are straightforward computations and have been used previously in research examining school effectiveness.

The school effectiveness indices associated with each of the study models are as follows:

1. Within-school least squares regression. This model allows schools to be tested for differential effectiveness at various combinations of predictor values. Since a single school effectiveness index may be misleading using this model, three seperate indices were calculated for each school. Mean individual pretest subtest scores across all schools and points one standard deviation above and below the mean served as reference points to represent low-, middle-, and high-scoring students. The regression estimates of the mean posttest composite scores at these three reference points served as school effectiveness indices.
2. Pooled least squares regression. In this model, the effectiveness index calculated for each school was simply the average of the residuals for the individuals within each school.
3. Pooled least squares regression with adjusted alphas. Since the regression coefficients for this model were assumed to be the same for all schools while the intercepts differed, a school effectiveness index was compiled by calculating regression estimates for a

specified set of reference points. For any given set of reference points the relative effectiveness of a school in relation to all other schools in the analysis can be obtained by comparing the resulting regression estimates. For this analysis the overall mean pretest subtest scores served as reference points.

4. Bayesian m-group regression with identical regression coefficients. Similar to the pooled least squares regression with adjusted alphas, a single school effectiveness index was calculated based on means representing middle-scoring students. This index represents regression estimates of the mean posttest composite score at reference points equivalent to the overall pretest subtest means.

5. Least squares regression using school mean scores. For this model, two effectiveness indices were calculated. The first index simply represented the school residual obtained by subtracting the observed mean posttest score from the predicted mean posttest score. In addition the following ratio suggested by Dyer, Linn and Patton (1967) was calculated for each school:

$$I = \frac{O - P}{\overline{SD} / \sqrt{\bar{n}}}$$

where O is the observed ouput mean for a particular school, P is the predicted output mean for a particular school, $\overline{SD}$ is the average within-school standard deviation for the output measure and $\bar{n}$ is the average number of students per school with output measures. A second school effectiveness index was then calculated using the following system:

$$I < -1.5,\ SEI = 1;$$
$$-1.5 < I < -.5,\ SEI = 2;$$
$$-.5 < I < .5,\ SEI = 3;$$
$$.5 < I < 1.5,\ SEI = 4;$$
$$1.5 < I,\ SEI = 5.$$

6. Mean difference scores. The computed mean difference score, that is, the mean posttest composite score minus the mean pretest composite score served as the school effectiveness index.

METHODOLOGY

Examining the Predictive Efficiency of Regression Models

There are several criteria that are commonly used for comparing predictive efficiency of regression models. The most common criteria is the mean squared error (MSE), which is the average of the squared differences between the predicted and observed measure within each school or across all subjects. Other commonly used criteria are the mean absolute error (MAE), which is the average of the absolute differences between predicted and observed, and the correlation (CORR) between the observed scores and the predicted scores.

Regression estimates were obtained for each school and the total sample using the Bayesian m-group regression model and the four classical regression models which use individual student data as input and output. Normal curve equivalent (NCE) scores for the five major subtests of the ITBS administered in 1976 to third graders served as predictor variables, while the composite NCE score of the ITBS administered in 1978 to fifth graders served as the criterion variable. The Bayesian model was compared with the classical regression models by calculating the mean squared error, mean absolute error, and the correlation between the observed and the predicted scores for each school.

Cross-Validation

In an attempt to provide more insight into the predictive efficiency of the regression models of this study a cross-validation analysis was undertaken. Generally, the correlations of predicted scores with observed scores are overestimates since the zero order correlations are assumed to be error free. The regression equation is fitting idiosyncrasies found in the sample and this imprecisely estimates the true relationship.

To estimate the degree of overestimation of R, a double cross-validation analysis was performed. For the first cross-validation, Cohort I, those students having complete sets of 1976 third grade ITBS scores and 1978 fifth grade ITBS scores, served as the screening sample. Estimates of the regression coefficients based on Cohort I students for each regression technique were then applied to Cohort II, the calibration sample. Cohort II consisted of those students having complete sets of 1977 third grade ITBS scores and 1979 fifth grade ITBS scores. A predicted output score as well as a residual score was calculated for each subject in Cohort II.

For the second cross-validation, Cohort II served as the screening sample and Cohort I served as the calibration sample. That is, regression coefficients obtained from Cohort II were applied to Cohort I. Predicted output scores and residual scores were calculated for each subject in Cohort I.

The predictive efficiency of the regression models was compared using three criteria: the mean squared error (MSE), mean absolute error (MAE), and the correlation (CORR) between the observed and the predicted scores.

Comparison of School Effectiveness Indices

A total of nine school effectiveness indices were derived from the six models: (1)-(3) regression estimates of within-school least squares regression lines at three sets of reference points (low, middle, high); (4) mean school residuals from pooled least square regressions; (5) regression estimates from pooled least square regressions with adjusted alpha for a middle set of reference points; (6) regression estimates from Bayesian m-group equal slopes regressions for middle reference points; (7) school residuals from least square regressions using school means; (8) classifications based on the Dyer, Linn, and Patton (1967) classification scheme; and (9) mean composite difference scores.

The intercorrelations among the school effectiveness indices derived for Cohort I were examined and direct comparisons made. In addition, indirect comparisons were obtained by examining the correlations of the school effectiveness indices with the school and community variables that did not enter into the regressions. The correlation patterns with other variables were examined as a means of providing a basis for interpreting the school effectiveness indices.

Estimating the Stability of School Effectiveness Indices Across Samples

To insure that differences in school effectiveness indices represent more than just error variation, the consistency of the effectiveness measures across random samples of students from the same grade level of the local educational system was examined. Estimates of this type of reliability concerned with students as a source of error were computed using two methods.

Students from Cohort I were randomly divided into halves within each school. Each of the six models were used to compute school effectiveness indices for each school in each sample. A total of nine school effectiveness indices were computed for each school for both samples. Correlations between these indices for the two random halves were computed and the results were compared.

A second reliability coefficient for the school effectiveness indices was computed by means of analysis of variance. Using the two random halves from each school, the variations of a particular set of school residuals were divided into among-school and among-sample variation. Winer (1962) indicates that the reliability of the mean of two observations is estimated by

$$\frac{\sigma^2_{\text{schools}}}{\sigma^2_{\text{schools}} + \sigma^2_{\text{samples}}/2}$$

In this case the observations were the school effectiveness indices for the two samples.

Estimating the Stability of School Effectiveness Indices Across Consecutive Classes

In addition to measuring the stability of indices across random samples it was also desirable to examine the consistency of school effectiveness indices for consecutive classes in the same school. This type of stability measure deals with students and factors which vary over time as sources of error. Assuming that no drastic out-of-the-ordinary changes occurred in the educational programs, school effectiveness indices should remain relatively stable for consecutive classes. If this holds true then schools with consistently high or low effectiveness indices should be identified and examined for possible causes of the observed results.

Matched-longitudinal samples, Cohort I and Cohort II, were formed for two time periods, 1976-78 and 1977-79, respectively. Nine effectiveness indices were computed for each school for both time periods using the six methods. Correlations between each set of indices for each school were then calculated and compared.

Chapter IV

RESULTS

SELECTION OF REGRESSION VARIABLES

Five major subtest scores and a composite score from the ITBS were available as achievement variables for entry into the regression analyses of this study. The means and standard deviations of these achievement scores for each school in Cohort I (Table 3 and Table 4) and Cohort II (Table 5 and Table 6) were calculated. In addition, numerous school and community variables were available as non-achievement variables for inclusion into the regression equations. No individual student socio-economic or family background information was available for this study. School and community information obtained from school records were tagged to individual students from their respective schools.

Intercorrelations between achievement variables and non-achievement variables were calculated for Cohort I (Table 7). Only the average pupil attendance (r=.25) and mobility rate (r=-.22) had correlations with the fifth grade composite variable that gave some promise for inclusion into the regression equations. Initial regression analyses revealed that the third grade achievement subtest scores were the first variables to enter as independent variables and that

Table 3

Means and Standard Deviations of ITBS Tests
For Cohort I (76-78) Grade 3

		Vocabulary		Reading		Language		Word Study		Arithmetic		Composite	
School	N	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
51	63	57.37	20.97	56.32	14.82	59.99	17.03	59.17	19.91	60.91	16.13	60.49	16.13
101	24	48.63	20.72	47.63	18.70	50.58	17.80	51.04	17.16	51.54	21.99	50.44	16.97
102	47	59.43	22.05	58.77	19.91	65.14	13.09	61.18	16.57	61.82	19.23	62.36	15.11
106	35	45.51	23.55	41.29	24.65	44.75	20.46	46.16	21.09	46.01	22.14	45.12	20.07
205	26	56.23	14.68	53.08	17.18	56.15	15.32	55.85	11.65	53.62	10.36	55.34	11.08
206	52	51.63	17.92	51.44	18.86	48.48	16.35	50.13	13.41	50.37	18.07	49.83	16.04
207	25	55.00	16.86	48.04	22.89	53.89	16.49	54.23	19.94	56.44	18.39	54.01	17.18
209	44	58.09	17.53	56.98	19.46	62.87	13.82	62.57	15.96	64.38	15.31	62.09	13.87
220	24	71.21	13.59	59.00	11.91	70.42	8.45	68.53	11.85	74.98	13.68	69.77	8.64
225	12	64.08	22.05	61.50	26.37	70.00	19.15	68.25	15.83	70.04	18.20	68.22	17.83
226	72	65.47	14.91	64.01	17.30	67.84	11.02	67.63	14.46	63.88	15.23	66.50	12.20
235	74	63.62	14.77	60.70	20.32	65.64	13.18	66.73	14.61	69.27	16.16	65.96	13.18
302	30	73.00	22.54	56.73	18.48	60.48	17.40	51.52	17.04	62.18	18.37	61.87	16.78
304	23	51.35	19.46	53.09	19.10	68.72	11.44	63.77	17.04	63.37	23.43	63.40	14.53
307	11	53.09	20.85	46.27	29.61	51.73	17.40	53.33	18.12	50.64	17.82	51.60	17.83
308	53	62.91	16.63	59.91	20.06	64.19	12.58	58.22	17.62	60.10	17.22	61.31	14.28
310	59	69.05	19.35	64.34	16.76	69.28	11.89	66.58	14.46	74.61	13.59	69.04	12.43
312	38	59.95	16.25	53.53	19.92	56.91	15.77	54.04	15.27	57.00	19.02	56.11	14.09
313	63	64.27	15.97	62.89	19.29	66.44	15.17	64.75	15.80	62.09	19.30	64.67	15.31
409	13	76.08	10.50	79.69	11.36	73.08	9.90	76.85	10.25	78.92	8.77	76.04	7.95
414	32	73.69	14.22	70.47	12.36	67.76	13.74	72.25	12.88	69.80	17.19	70.14	11.89
415	24	69.86	19.32	65.71	22.58	67.49	12.22	68.78	16.24	69.15	14.85	68.20	13.38
416	19	68.16	22.39	67.21	23.05	57.09	14.82	71.88	14.38	72.37	18.22	69.46	15.13
422	32	75.16	14.67	77.69	17.46	74.63	12.54	75.70	13.45	77.45	11.91	75.76	11.64
504	46	68.11	16.65	68.80	18.85	74.20	11.74	72.99	14.99	72.60	16.19	72.53	13.29
505	63	69.78	15.01	67.13	16.27	68.42	12.49	71.16	13.70	68.39	16.13	69.17	12.66
509	41	64.46	14.32	55.12	15.78	61.48	11.26	63.51	13.36	62.65	21.01	61.94	12.35
552	45	55.56	17.90	53.07	25.30	59.61	18.50	55.44	20.38	56.54	22.91	56.95	18.99
553	38	37.03	15.54	35.68	18.01	40.88	15.37	41.87	12.53	38.21	15.18	39.84	12.15
559	32	53.72	17.16	52.69	23.28	58.45	17.78	58.33	18.74	59.42	21.13	57.54	18.11
561	37	67.49	16.91	63.54	17.45	63.00	12.55	66.72	15.43	68.85	14.41	65.54	12.87
563	37	60.14	16.11	57.54	23.19	62.50	15.09	58.57	17.63	56.64	15.05	59.70	14.86
566	25	49.40	21.33	49.96	22.90	49.02	16.62	46.75	22.25	46.12	20.56	47.99	18.45
570	53	64.34	13.36	66.30	17.89	72.58	10.20	67.61	13.71	72.50	15.64	69.89	11.27
601	65	79.17	17.88	65.63	19.95	72.00	13.45	76.51	14.59	21.47	19.07	73.21	14.05
604	41	70.56	15.42	65.00	16.29	70.77	11.06	67.69	11.64	68.88	15.88	69.04	10.78
652	26	52.73	17.88	51.04	17.99	50.68	15.29	50.78	18.86	46.31	16.33	50.13	15.18
767	33	62.00	18.55	58.67	19.62	61.84	16.51	60.44	16.90	60.70	19.85	60.98	15.65
768	18	55.11	17.53	54.83	19.34	56.43	18.00	59.26	15.91	52.50	18.23	56.22	16.46
769	53	59.77	16.65	58.72	19.25	60.55	12.73	60.65	15.87	55.32	18.54	59.39	14.50
772	43	42.42	19.21	43.37	19.12	47.95	15.93	43.64	16.07	42.63	18.59	44.89	15.16
774	44	53.98	16.81	50.73	21.53	58.01	17.80	49.59	20.21	55.17	19.87	54.17	17.27
785	32	57.69	16.35	57.09	17.94	63.95	16.65	54.83	18.27	63.52	21.86	60.19	16.86
786	25	56.04	15.79	54.56	20.87	59.31	15.43	59.19	16.34	65.12	14.75	59.79	14.67
788	65	56.42	17.43	53.95	19.30	57.23	16.23	54.22	16.48	51.85	20.11	55.06	15.80
795	35	57.77	19.74	60.77	20.00	59.68	15.02	59.31	16.56	61.70	19.79	59.87	15.99
797	35	57.80	17.32	52.31	21.78	62.67	14.43	61.49	18.10	63.20	15.30	61.06	14.54
808	35	63.06	14.50	64.63	17.94	60.39	16.40	60.99	15.83	58.11	15.59	60.77	14.24
821	52	62.71	22.26	62.06	23.75	65.27	14.17	65.83	19.92	65.69	18.99	64.97	16.47
822	32	56.59	24.95	57.59	21.34	55.86	17.35	55.19	20.94	57.22	21.05	56.15	18.57

Table 4

Means and Standard Deviations of ITBS Tests
For Cohort I (76-78) Grade 6

		Vocabulary		Reading		Language		Word Study		Arithmetic		Composite	
School	N	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
51	63	57.95	21.21	55.37	20.27	58.15	15.78	59.56	18.48	60.38	19.52	58.45	17.22
101	24	47.29	17.50	45.83	16.24	50.11	14.72	49.15	17.11	51.58	21.65	49.25	15.35
102	47	67.30	21.18	54.64	20.00	58.60	16.46	60.49	18.38	63.79	20.88	60.31	17.14
106	35	46.11	21.42	42.49	22.23	47.51	18.74	49.56	20.63	47.71	21.69	47.15	19.00
205	26	53.00	12.09	51.50	12.43	54.73	13.39	54.62	12.80	53.37	11.12	53.39	11.34
206	52	45.46	20.90	45.81	21.21	46.25	16.63	49.84	17.78	50.82	19.93	47.55	16.79
207	25	50.76	25.70	52.76	24.53	51.42	17.94	52.95	21.87	56.34	20.27	52.39	19.26
209	44	59.41	18.60	59.39	20.70	63.45	13.39	64.81	15.74	68.59	17.15	63.81	14.22
220	24	66.67	9.28	58.21	10.11	64.76	6.84	61.03	13.98	69.00	15.36	63.51	9.18
225	12	59.25	20.12	51.50	21.35	59.79	13.85	61.72	14.42	65.50	15.91	59.63	14.29
226	72	62.78	15.17	62.76	17.02	63.13	12.95	68.25	14.71	69.65	16.73	65.79	12.87
235	74	62.07	15.95	59.89	17.65	62.81	12.70	66.17	17.33	71.30	17.29	64.62	13.49
302	30	56.70	17.13	51.63	16.21	50.35	16.38	58.68	14.66	66.73	17.19	55.72	15.21
304	23	50.00	14.43	44.52	18.21	58.73	13.84	52.45	18.48	57.11	23.60	53.24	15.11
307	11	45.91	18.52	44.09	22.77	49.84	18.32	49.42	20.79	47.95	16.99	48.77	18.37
308	53	58.08	17.84	60.23	19.02	67.87	12.26	56.94	15.71	65.97	17.57	65.91	13.96
310	59	60.80	15.57	59.51	17.93	64.39	12.89	65.58	15.80	70.19	15.49	64.52	13.49
312	38	61.26	14.22	61.26	15.30	59.13	15.75	63.53	16.69	61.17	21.02	62.27	14.92
313	63	62.25	15.39	61.89	18.01	66.68	14.17	73.24	15.30	69.41	19.51	68.19	14.68
409	13	65.69	9.37	66.31	17.06	67.23	13.71	69.00	14.53	75.27	10.08	67.33	11.86
414	32	69.38	13.70	71.34	12.01	68.27	14.31	72.41	13.58	75.48	17.52	70.76	12.91
415	24	73.04	18.51	72.33	18.80	72.44	12.47	74.62	15.91	76.44	15.76	74.17	13.97
416	19	70.16	10.80	70.00	17.10	68.43	11.92	74.33	17.21	71.79	16.90	70.29	12.93
422	32	75.50	13.39	78.13	13.97	77.19	11.02	81.10	10.55	81.52	12.41	78.96	9.81
504	46	63.85	15.54	63.24	16.33	62.58	14.54	65.98	16.28	72.57	16.90	64.45	14.29
505	63	67.11	15.41	66.02	16.10	67.99	13.62	71.58	15.96	72.55	18.41	69.58	14.07
509	41	57.05	16.91	54.37	16.46	59.28	15.61	53.63	16.15	64.15	19.96	60.24	14.99
552	45	53.07	19.32	50.93	17.95	52.47	16.72	52.41	18.81	55.79	22.37	52.53	16.92
553	38	34.37	17.93	39.32	15.42	39.40	16.18	42.41	13.74	40.39	15.45	39.77	13.84
559	32	55.09	13.89	53.56	17.66	58.20	16.37	62.24	16.94	67.22	18.44	60.01	14.91
561	37	62.62	14.35	59.95	17.42	58.58	13.22	59.60	15.25	68.76	14.93	60.08	13.18
563	37	57.19	17.75	55.73	19.48	57.28	15.15	58.79	17.65	57.54	17.71	57.32	15.73
566	25	47.36	20.01	45.08	21.00	50.12	16.51	49.19	17.22	46.74	21.14	48.59	17.14
570	53	64.79	15.36	61.98	16.79	64.51	14.28	66.04	14.60	67.13	16.77	64.11	13.81
601	65	67.83	16.51	65.23	16.12	66.74	13.23	69.93	16.39	69.92	18.29	67.37	14.47
604	41	68.07	16.92	65.93	14.77	66.41	11.55	66.74	12.37	69.84	15.76	66.85	11.23
652	26	49.15	17.54	49.27	17.36	53.68	13.40	51.17	13.97	46.79	17.73	51.04	13.67
767	33	59.03	13.84	60.58	15.59	59.29	15.25	65.89	16.27	67.74	17.30	62.91	13.85
768	18	54.44	19.64	55.33	22.65	52.94	19.86	57.91	17.30	56.06	19.84	55.30	18.40
769	53	55.51	18.65	57.17	17.06	53.01	13.21	52.84	18.82	56.45	18.68	53.85	15.32
772	43	43.79	17.22	47.28	14.34	54.02	12.36	53.26	13.39	46.48	14.87	51.40	12.33
774	44	61.02	21.34	54.73	19.83	60.68	18.48	62.75	19.28	64.41	18.15	61.89	17.38
785	32	58.63	16.58	55.44	19.53	59.95	13.58	63.96	16.73	64.02	20.04	60.92	14.81
786	25	60.56	15.19	58.96	18.13	67.67	13.08	68.08	16.09	67.00	17.11	65.99	13.71
788	65	51.32	18.94	50.65	20.38	55.19	15.94	54.99	16.53	55.19	18.97	54.46	15.61
795	35	59.31	16.16	58.09	16.13	58.88	15.85	61.39	16.94	60.59	20.13	59.65	15.65
797	35	66.51	24.00	54.14	19.02	61.97	17.38	68.57	15.39	66.04	15.50	64.36	15.91
803	35	64.20	16.51	63.66	15.45	66.30	12.16	64.77	15.27	62.71	18.81	65.45	13.21
821	52	67.00	18.48	62.77	22.35	63.99	14.82	69.34	17.10	67.49	19.76	65.87	16.21
822	32	54.13	23.17	57.56	19.32	56.91	17.55	61.66	18.30	59.80	24.98	58.85	18.35
TOTAL	1946	59.08	19.01	57.52	19.30	60.03	16.12	62.33	18.01	63.45	20.00	60.76	16.42

Table 5

Means and Standard Deviations of ITBS Tests
For Cohort II (77-79) Grade 3

		Vocabulary		Reading		Language		Word Study		Arithmetic		Composite	
School	N	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
51	60	57.40	21.36	57.15	20.86	62.77	15.20	59.19	15.16	64.47	16.91	61.91	15.08
101	23	55.13	16.98	57.22	18.20	65.00	11.85	62.20	14.29	57.83	13.76	61.33	11.96
102	33	72.15	16.53	68.09	14.73	76.41	12.27	73.63	15.57	75.47	17.45	74.34	13.57
106	55	56.11	21.68	54.95	20.90	63.52	18.27	56.10	20.19	56.87	23.01	58.83	18.57
205	31	58.26	14.31	51.65	18.12	65.52	10.53	64.41	11.84	51.71	16.08	60.78	11.35
206	39	51.56	18.81	56.67	16.86	56.74	16.69	53.62	20.58	52.46	20.92	54.63	17.31
207	38	58.21	17.51	62.05	18.13	64.99	17.18	63.54	17.03	62.72	15.96	63.30	15.64
209	40	60.20	12.33	60.72	17.20	67.63	12.92	66.50	13.09	60.61	17.31	64.77	12.82
220	44	68.43	12.47	66.48	13.85	71.69	10.90	70.64	11.06	70.09	11.14	70.35	9.63
225	25	61.22	14.14	56.04	15.08	67.12	10.59	61.13	13.39	57.15	12.69	62.13	10.46
226	56	62.55	18.77	60.71	20.03	63.58	14.55	61.90	15.33	59.53	18.37	62.03	14.66
235	57	65.07	13.98	62.91	17.11	66.26	13.70	65.36	12.99	66.19	16.25	65.59	12.08
302	35	72.97	21.21	59.71	20.80	64.61	15.91	60.78	18.56	63.97	20.06	63.77	17.06
304	15	40.27	15.80	30.07	20.29	42.95	19.35	41.11	18.13	46.23	20.33	41.63	16.85
307	16	53.00	15.97	45.31	19.88	58.08	14.87	55.90	17.44	50.78	23.28	54.53	15.77
308	49	64.14	14.62	60.55	17.48	63.21	13.38	62.51	15.33	61.59	16.71	62.57	13.52
310	66	63.30	18.84	62.97	22.36	70.65	15.68	64.41	18.92	68.32	18.68	67.16	16.78
312	42	61.19	18.37	64.57	18.31	68.45	13.31	65.67	16.75	68.23	13.83	66.64	14.01
313	37	61.81	11.41	61.51	15.30	63.67	12.48	62.86	14.81	61.14	15.08	62.62	12.23
409	13	66.92	13.69	63.85	13.99	63.56	10.07	58.56	17.58	52.81	21.07	60.57	13.48
414	38	73.58	16.58	70.47	17.88	76.11	9.90	70.26	14.92	74.03	14.94	73.39	11.92
415	28	71.64	12.45	71.57	15.76	74.61	11.73	71.92	12.02	73.32	13.97	73.09	10.53
416	12	57.58	15.50	52.08	20.30	60.23	13.94	54.14	16.89	55.38	16.63	56.70	14.91
422	53	70.53	14.78	71.00	17.90	73.93	11.87	72.48	13.41	72.81	12.87	72.75	12.00
504	37	70.76	13.87	71.11	17.78	77.51	9.34	75.41	12.94	74.28	14.12	75.15	10.93
505	59	67.63	13.92	65.36	15.67	71.58	12.17	70.34	12.41	68.69	14.45	69.84	11.52
509	18	68.22	12.99	67.00	13.28	73.43	10.57	73.46	11.29	71.19	13.80	71.97	9.86
552	43	57.86	21.64	61.65	20.79	62.47	16.24	59.19	21.16	59.33	19.68	60.51	17.84
553	37	50.89	17.24	53.03	19.64	52.84	17.89	52.91	18.44	53.11	21.32	52.75	17.07
559	33	63.15	16.40	69.27	13.00	71.71	13.66	66.95	17.01	69.39	19.13	68.99	14.19
561	58	59.64	15.07	59.45	16.97	63.81	15.36	59.80	19.08	63.46	17.26	61.88	15.18
563	45	59.67	17.95	60.49	18.70	69.26	14.54	63.44	16.88	71.78	15.59	66.46	14.54
566	15	54.27	23.87	56.27	23.74	58.58	19.32	56.78	18.83	53.63	19.45	56.59	18.32
570	62	65.47	18.16	62.26	19.92	72.63	14.13	69.60	14.84	70.01	15.62	69.73	14.09
601	68	81.56	17.14	74.01	16.71	78.90	13.63	78.73	14.79	76.54	16.76	78.22	13.88
604	36	71.28	14.84	70.19	14.98	73.74	9.67	70.97	13.33	65.04	14.66	70.86	10.65
652	27	59.00	18.56	57.00	24.29	60.82	15.63	60.48	17.72	56.43	19.57	59.42	16.33
767	35	56.77	17.59	55.60	20.20	59.69	14.87	59.65	19.24	61.24	16.08	59.32	15.61
768	22	64.95	17.52	65.68	17.69	70.06	14.48	69.83	17.79	72.91	15.35	69.65	14.79
769	38	55.74	15.24	55.53	18.53	59.63	13.90	62.32	14.84	57.96	17.40	59.33	13.65
772	52	55.94	23.28	57.65	21.71	62.14	18.10	60.24	20.49	57.49	20.75	59.81	18.65
774	45	55.78	16.11	51.07	20.63	57.78	16.65	52.33	18.90	57.53	21.48	55.46	17.00
785	38	55.71	18.70	53.58	21.06	58.48	19.11	56.37	20.84	58.72	16.41	57.25	17.68
786	25	58.16	15.74	56.92	17.68	64.32	13.09	57.01	16.68	57.16	16.32	59.79	13.98
788	47	51.57	21.24	47.11	22.55	53.32	15.35	55.65	18.78	52.11	21.18	53.01	17.34
795	30	69.90	19.84	68.13	20.68	72.03	14.57	66.26	19.07	69.37	19.61	69.42	16.94
797	50	60.92	15.15	59.92	17.73	66.38	14.90	66.45	15.09	68.38	17.55	65.68	14.34
808	28	62.00	19.47	65.79	23.21	67.68	13.30	63.98	15.81	62.91	17.63	65.11	15.29
821	45	66.20	15.89	62.29	17.78	63.59	12.37	66.34	15.26	64.32	17.89	64.73	13.51
822	37	59.32	14.79	59.00	17.38	61.55	10.93	59.59	12.29	56.32	14.99	59.66	11.59
TOTAL	1933	62.36	18.45	61.06	19.70	66.15	15.62	63.80	17.49	63.71	18.58	64.26	15.31

Table 6

Means and Standard Deviations of ITBS Tests
For Cohort II (77-79) Grade 5

		Vocabulary		Reading		Language		Word Study		Arithmetic		Composite	
School	N	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
51	60	55.22	19.97	54.78	22.23	56.24	15.81	58.86	17.27	59.83	20.22	57.38	16.82
101	23	52.65	20.61	54.96	18.92	54.10	15.34	56.49	16.28	51.46	17.81	54.22	15.62
102	33	78.82	17.52	62.97	14.93	64.76	10.42	65.91	13.35	66.62	14.25	66.53	11.64
106	55	50.82	22.06	52.45	22.70	56.20	18.89	55.00	19.14	55.05	20.87	54.63	18.61
205	31	50.13	17.36	48.39	15.53	54.96	13.98	54.68	14.90	48.02	18.51	52.58	13.70
206	39	48.72	19.96	52.79	20.45	51.59	14.65	53.74	16.14	57.46	18.85	53.09	15.29
207	38	60.76	18.89	56.21	20.85	59.18	17.54	64.23	19.94	66.33	19.72	61.73	17.76
209	40	59.07	14.33	56.02	16.70	58.79	14.65	60.82	18.04	55.75	19.97	58.57	15.03
220	44	65.48	11.99	66.41	12.84	77.40	8.85	74.32	12.14	74.01	14.09	73.86	9.93
225	23	51.00	13.57	47.04	16.83	52.76	12.70	49.51	12.65	44.28	15.78	49.68	12.00
226	56	61.91	13.92	61.02	15.67	64.85	12.47	63.57	14.74	63.62	16.14	63.66	12.35
235	57	64.49	14.58	52.93	15.27	62.91	12.93	64.98	13.13	64.99	18.09	54.00	12.74
302	35	55.37	21.83	56.66	19.61	57.09	17.51	55.58	17.60	59.94	20.43	57.27	17.31
304	15	41.20	16.61	41.33	11.23	48.05	16.76	44.02	17.72	44.13	16.96	45.01	15.08
307	16	46.69	23.07	42.63	19.79	48.75	18.50	51.58	21.34	51.53	21.33	49.28	18.81
308	49	64.55	13.72	65.12	17.65	69.84	9.56	66.65	14.19	68.99	16.62	67.91	11.81
310	66	62.39	17.45	50.80	19.12	64.57	14.94	66.69	18.13	61.83	22.00	64.11	16.41
312	42	61.81	18.32	60.26	20.33	65.32	12.46	65.26	14.96	65.48	15.85	64.55	13.64
313	37	60.65	13.97	63.68	15.13	66.72	10.59	72.63	11.61	68.78	13.64	67.88	10.67
409	13	66.54	18.50	66.15	15.74	61.79	12.07	63.85	22.54	62.92	19.96	63.38	16.25
414	38	70.08	15.72	70.13	17.93	73.76	10.56	74.21	13.77	78.74	12.89	74.12	11.13
415	28	71.39	15.12	82.68	15.56	78.77	9.28	73.74	11.19	77.30	14.49	76.81	9.49
416	12	54.08	16.90	56.50	10.62	58.04	11.05	54.50	18.24	54.38	14.07	55.91	12.88
422	53	70.96	14.71	68.81	14.96	71.70	12.75	71.10	15.84	72.42	15.11	71.34	12.86
504	37	67.73	17.04	64.41	18.64	66.05	12.74	67.84	15.69	70.36	16.36	67.33	13.55
505	59	63.00	13.52	61.03	15.05	6[illegible].73	12.24	63.69	16.52	66.69	17.76	63.95	13.20
509	18	65.17	14.00	62.72	17.03	70.58	10.84	71.54	14.21	66.31	18.70	68.86	12.39
552	43	55.79	19.66	55.26	20.11	55.84	14.77	55.15	18.94	54.90	19.76	55.42	16.50
553	27	47.32	22.13	51.32	18.60	54.47	17.16	54.82	16.33	53.07	24.56	53.37	17.58
559	33	63.61	15.33	65.33	15.60	67.27	15.58	69.93	17.41	73.42	18.64	68.61	15.33
561	58	57.57	21.41	51.90	21.41	55.23	15.75	59.06	13.40	62.09	13.69	57.43	16.32
563	45	56.56	17.06	53.73	20.66	58.85	14.79	60.62	19.33	62.01	20.10	59.23	16.21
566	15	54.87	18.47	53.87	20.65	54.28	21.07	59.02	16.79	52.07	18.24	55.19	17.98
570	62	59.95	16.78	59.53	18.69	64.59	13.66	66.63	15.00	61.80	17.30	63.76	14.11
601	68	72.71	15.06	67.56	16.73	70.64	11.78	74.69	13.50	76.29	15.02	72.68	11.90
604	36	70.97	15.18	64.67	15.24	66.85	9.70	69.04	14.51	69.07	17.78	68.03	11.86
652	27	57.37	22.10	54.96	21.83	59.75	12.47	57.52	17.40	57.33	18.70	58.05	15.47
767	35	53.36	19.37	54.83	15.63	52.34	17.40	54.99	19.18	55.49	16.79	54.00	16.60
768	22	61.73	21.12	61.91	19.74	60.20	16.60	67.95	18.86	64.45	21.17	63.38	17.46
769	38	55.26	14.62	51.11	19.02	54.19	12.80	56.61	13.98	58.38	14.28	55.43	12.16
772	52	55.98	21.49	56.67	22.45	62.51	16.58	62.78	17.41	60.01	18.33	61.00	16.66
774	45	53.04	23.87	47.76	21.29	53.44	19.39	53.01	21.83	56.26	23.87	53.28	20.06
785	38	53.32	18.79	54.42	18.14	52.66	17.07	54.51	16.60	51.99	16.07	53.26	15.49
786	25	67.00	16.53	62.32	15.94	58.24	9.65	68.52	11.78	69.38	15.58	67.87	11.12
788	47	49.40	19.95	48.21	19.66	51.13	14.35	52.15	16.79	50.61	20.07	50.89	15.56
795	30	67.97	19.14	64.87	22.94	64.65	16.05	64.52	18.34	62.12	23.07	64.48	17.86
797	50	58.04	16.90	58.30	20.42	61.38	15.37	66.67	18.61	63.70	18.60	62.66	16.02
808	28	60.43	18.10	61.04	20.01	63.62	14.92	65.29	16.60	62.05	18.24	63.26	15.53
821	45	63.87	15.26	60.91	18.00	59.58	14.12	64.78	15.36	65.09	21.71	62.51	14.81
822	37	55.24	16.53	54.35	13.80	57.20	12.63	57.24	13.46	52.97	15.07	56.01	12.31
TOTAL	1933	59.98	18.93	58.67	19.42	61.50	15.70	62.70	17.61	62.35	19.64	61.59	16.12

no school or community variable contributed significantly to the prediction of the fifth grade total composite score.

It was therefore decided to restrict the regression analyses to achievement variables only, something that is commonly done in school effectiveness research, and use the five third grade NCE subtest scores as independent variables and the fifth grade NCE composite variable as the dependent variable. The school and community variables were examined later in the study as correlates to school effectiveness indices. The correlational pattern of these variables with the school effectiveness indices provided a basis for comparing and interpreting these indices.

COMPARISON OF THE REGRESSION MODELS

In Table 8 the estimated regression coefficients for the four regression models using individual scores are provided for each of the 50 schools for Cohort I. The symbol α corresponds to the sample intercept estimate while β_1, β_2, β_3, β_4, and β_5 refer to the five regression coefficients corresponding to the five ITBS subtest scores. The regression coefficient estimates for the five ITBS subtest scores obtained from the pooled least squares regression, the pooled least squares regression with adjusted alpha, and the Bayesian m-group equal slope model are located at the bottom of the table.

Table 7

Intercorrelations of ITBS Scores and Non-Achievement Variables
For Cohort I (76-78)

	ITBS Subtest and Composite Scores					
Non-Achievement Variables	Vocabulary	Reading	Language	Word Study	Arithmetic	Composite
Average Pupil Attendance	0.2038	0.1757	0.2234	0.2034	0.2062	0.2470
Mobility Rate	-0.2209	-0.1681	-0.2067	-0.2071	-0.2013	-0.2230
Pupils Transported	0.0649	0.0157	0.0720	0.0675	0.0656	0.0277
Average Class Size	0.0475	0.0743	0.1441	0.1041	0.1005	0.1014
Student/Professional Staff Ratio	0.1043	0.0951	0.1825	0.1333	0.1507	0.1506
Total Student Enrollment	-0.1313	-0.0856	-0.0559	0.1084	-0.0827	-0.0994
Percent of Minority Students	-0.0992	-0.0824	-0.0642	-0.0978	-0.0555	-0.0595
Staff Inexperience	-0.0137	-0.0108	-0.0492	-0.0002	-0.0093	-0.0393
Percent of Minority Staff	-0.0571	-0.0582	-0.1041	-0.0887	-0.0840	-0.0694
Percent of Male Staff	-0.0366	-0.0442	-0.0622	-0.0861	-0.0203	-0.0274
Property Destruction	-0.0216	0.0117	0.0820	0.0470	0.0359	-0.0105
Special Education Availability	-0.0315	0.0012	-0.0404	-0.0082	-0.0342	-0.0465

Note. ITBS Subtest Scores from Grade 3 and Composite Score from Grade 5.

Table 3

Regression Estimates for Cohort I (76-78)

Schools	Within LSQ α	β_1	β_2	β_3	β_4	β_5	Pooled LSQ with Adjusted α α	Bayesian Equal Slope Model α
51	-1.647	-0.017	0.232	0.467	0.245	0.064	6.665	6.770
101	8.716	0.183	0.084	0.090	0.283	0.168	5.809	6.018
102	-5.199	0.062	0.163	0.481	0.087	0.253	6.317	6.926
106	8.624	0.079	0.234	0.183	0.383	-0.013	8.281	8.310
205	13.294	0.104	0.009	0.328	0.322	-0.039	6.210	6.384
206	1.477	0.170	0.086	0.192	0.162	0.307	4.160	4.345
207	-4.727	-0.098	-0.004	0.601	0.141	0.401	5.868	6.074
209	5.415	0.158	0.123	0.315	0.029	0.320	10.502	10.456
220	7.320	0.075	0.182	0.037	0.445	0.094	3.552	3.959
225	15.702	0.626	-0.114	-0.818	0.602	0.385	1.182	2.196
226	0.279	0.103	-0.020	0.418	0.264	0.216	8.603	8.658
235	6.195	0.149	0.103	0.035	0.405	0.194	7.812	7.888
302	8.663	-0.062	0.379	0.670	-0.049	-0.119	1.955	2.389
304	6.384	0.379	-0.026	0.150	-0.028	0.320	-0.231	0.487
307	10.325	-0.321	0.415	0.049	0.502	0.137	4.452	5.021
308	8.156	0.140	0.096	0.338	0.031	0.328	13.010	12.884
310	-2.228	0.079	0.232	0.239	0.200	0.221	4.809	4.979
312	16.269	0.269	0.251	0.007	-0.035	0.315	13.686	13.452
313	15.494	0.290	0.230	0.065	0.094	0.149	12.537	12.464
409	-15.295	0.152	-0.009	0.725	-0.680	0.901	1.020	1.986
414	-1.163	0.086	0.264	0.293	0.144	0.240	9.811	9.765
415	13.734	0.378	0.027	0.122	0.046	0.301	15.166	14.673
416	12.283	0.049	-0.073	-0.003	0.528	0.301	10.157	10.007
422	24.429	0.125	0.148	0.248	-0.035	0.230	13.157	12.915
504	-6.803	0.046	0.202	0.375	0.184	0.178	2.158	2.472
505	0.697	0.188	0.015	0.321	0.205	0.266	9.830	9.845
509	28.141	-0.010	0.300	-0.040	0.233	0.061	6.928	7.043
552	6.803	0.003	-0.098	0.248	0.247	0.395	3.657	3.897
553	0.744	0.081	0.111	0.433	0.201	0.156	5.800	5.953
559	21.100	0.488	0.109	-0.200	0.043	0.271	10.560	10.466
561	4.556	-0.017	0.168	0.380	0.278	0.051	3.153	3.461
563	-4.652	0.138	0.135	0.290	0.042	0.448	6.009	6.165
566	11.465	-0.019	0.248	-0.118	0.291	0.387	6.967	7.075
570	-3.869	0.329	0.187	0.173	0.325	-0.002	4.060	4.269
601	5.941	-0.229	0.215	0.398	0.470	0.012	4.298	4.493
604	3.026	0.243	0.101	0.128	0.347	0.110	7.354	7.449
652	9.593	0.103	0.130	0.108	0.280	0.210	7.671	7.735
767	20.405	0.150	0.102	0.282	0.128	0.035	10.275	10.200
768	-6.945	0.425	0.050	0.076	0.385	0.171	6.963	7.108
769	0.973	-0.157	0.388	0.207	0.241	0.222	2.701	2.953
772	20.325	0.215	-0.119	0.456	0.056	0.066	12.960	12.783
774	10.983	0.175	0.113	0.300	-0.042	0.369	15.313	15.045
785	19.069	0.198	0.269	-0.064	0.152	0.171	9.092	9.076
786	25.935	0.429	0.373	-0.196	0.194	-0.078	14.390	13.969
788	3.950	0.135	0.175	0.360	0.134	0.108	7.033	7.118
795	7.810	-0.172	0.132	0.400	0.282	0.213	7.755	7.817
797	-4.737	0.260	-0.213	0.548	0.344	0.155	12.188	12.027
808	15.715	-0.047	0.120	0.147	0.342	0.262	12.595	12.396
821	11.722	0.151	0.130	0.102	0.227	0.228	9.900	9.894
822	7.016	0.166	0.156	0.463	-0.158	0.284	10.065	9.987

Pooled LSQ	α=8.770	β_1=0.106	β_2=0.158	β_3=0.235	β_4=0.169	β_5=0.185
Pooled LSQ with Adjusted α		β_1=0.112	β_2=0.135	β_3=0.236	β_4=0.200	β_5=0.183
Bayesian Equal Slope Model		β_1=0.111	β_2=0.136	β_3=0.236	β_4=0.199	β_5=0.183

It is interesting to note the number of least squares regression weights for each of the five ITBS subtests that are negative. One might ask if these estimates are accurate or are they merely reflecting idiosyncrasies found in the data. It seems entirely plausable that the correct values for these coefficients are all zero or slightly positive. Examining the estimated regression coefficients for the remaining models would suggest that the negative values may be due solely to sampling variation.

The estimated regression coefficients obtained by the Bayesian m-group equal slope model are almost identical to those obtained using the pooled least squares regression with adjusted alpha. The pooled least squares regression coefficient estimates for the Language (β_3) and Arithmetic (β_5) subtests, and to a lesser extent, the Vocabulary (β_1) subtest, are very similar to those obtained from the Bayesian m-group equal slope model and the pooled least squares with adjusted alpha model. The regression coefficient estimate for the Reading (β_2) subtest is slightly higher while the regression coefficient for the Word Study (β_4) subtest is slightly lower for the pooled least squares model when compared to the Bayesian m-group equal slope model.

Examination of mean squared errors (Table 9), mean absolute errors (Table 10), and the correlations between the observed and predicted composite scores (Table 11) reveal no surprises. The within-school least squares regression minimized the error for each school. Since the Bayesian m-group equal slope regression represents a weighted average of a particular school's classical regression line and the regression line obtained for all schools collectively, it is not surprising that the mean squared errors, the mean absolute errors, and the correlations fall somewhere between those obtained from the within least squares regression and the pooled least squares regression. For Cohort I the average mean squared errors ranged from 40.44 for the within-school least squares regression to 63.38 for the pooled least squares regression, with the average mean squared error for the Bayesian approach and the pooled least squares regression with adjusted alpha between these extremes at 49.98 and 49.93, respectively. Similar patterns resulted for the mean absolute errors and the correlations between observed and predicted composite scores.

If residuals obtained from these regression equations are to be used as components of school effectiveness indices it is essential that they provide a fairly accurate estimate of the true relationship between subtest scores and total composite scores. It is entirely possible that the regression equations from one sample imprecisely estimates the true re-

Table 9

Mean Square Errors for Cohort I (76-78)

School	N	Within LSQ	Pooled LSQ	Pooled LSQ with Adjusted α	Bayesian Equal Slope Model
51	63	39.489	47.818	46.240	46.255
101	24	34.966	43.011	37.729	37.743
102	47	45.019	54.207	52.697	52.760
106	35	39.746	44.359	43.956	43.959
205	26	48.541	59.499	55.313	55.350
206	52	60.769	79.341	63.208	63.278
207	25	55.721	79.978	74.564	74.696
209	44	28.965	39.199	32.463	32.489
220	24	19.496	43.399	25.600	25.718
225	12	49.643	143.265	98.252	99.084
226	72	27.766	33.814	32.026	32.095
235	74	30.012	36.670	35.420	35.484
302	30	37.059	90.412	58.110	58.074
304	23	22.138	98.125	33.406	33.869
307	11	31.903	55.558	44.091	44.323
308	53	24.089	53.054	28.609	28.646
310	59	24.746	36.950	27.334	27.373
312	38	58.848	103.860	73.212	73.247
313	63	38.702	65.989	45.473	45.432
409	13	35.314	133.333	86.636	87.433
414	32	18.538	27.281	23.160	23.210
415	24	24.996	86.150	31.988	32.335
416	19	17.836	64.195	55.560	55.689
422	32	23.546	56.387	29.114	29.151
504	46	24.326	58.578	26.854	26.940
505	63	34.654	43.610	38.644	38.718
509	41	160.121	182.568	183.212	183.041
552	45	43.770	77.484	57.804	57.834
553	38	39.619	51.035	45.005	45.066
559	32	41.244	70.296	64.337	64.241
561	37	44.129	69.563	47.097	47.152
563	37	24.417	36.805	32.609	32.669
566	25	24.289	36.421	34.981	34.960
570	53	36.913	64.368	49.331	49.374
601	65	60.060	93.524	82.804	82.762
604	41	17.666	23.928	22.862	22.909
652	26	17.755	19.379	18.977	18.974
767	33	65.609	81.316	76.432	76.349
768	18	13.454	32.839	30.710	30.860
769	53	40.099	78.316	51.261	51.272
772	43	38.376	83.523	61.275	61.237
774	44	85.377	143.247	92.314	92.411
785	32	62.189	74.611	76.115	75.934
786	25	20.498	80.196	39.656	39.822
788	65	35.894	39.141	38.108	38.111
795	35	58.285	67.784	67.556	67.523
797	35	46.908	78.338	57.433	57.586
808	35	20.495	45.886	24.924	25.006
821	52	33.219	39.054	34.455	34.483
822	32	50.387	64.056	61.096	61.087
TOTAL	1946	40.442	63.378	49.926	49.973

Table 10

Mean Absolute Errors for Cohort I (76-78)

School	N	Within LSQ	Pooled LSQ	Pooled LSQ with Adjusted α	Bayesian Equal Slope Model
51	63	5.051	5.377	5.489	5.487
101	24	4.679	5.225	5.125	5.126
102	47	5.460	6.064	6.064	6.067
106	35	4.897	5.463	5.354	5.364
205	26	5.642	6.455	6.259	6.258
206	52	6.063	7.108	6.285	6.287
207	25	5.655	6.872	6.499	6.513
209	44	4.351	5.225	4.653	4.679
220	24	3.588	4.911	4.017	3.970
225	12	5.970	10.122	8.792	8.793
226	72	3.996	4.384	4.168	4.174
235	74	4.279	4.909	4.814	4.819
302	30	5.081	8.186	6.687	6.741
304	23	3.776	8.008	5.252	5.236
307	11	4.928	5.693	5.115	5.167
308	53	3.727	6.134	4.084	4.119
310	59	4.006	4.684	4.161	4.167
312	38	5.888	8.080	6.294	6.273
313	63	4.885	6.573	5.363	5.353
409	13	4.812	8.755	7.409	7.077
414	32	3.629	4.522	3.859	3.891
415	24	3.996	7.995	4.481	4.514
416	19	3.164	4.668	4.719	4.654
422	32	3.876	6.432	4.073	4.092
504	46	4.179	6.455	4.545	4.549
509	41	9.217	9.397	9.500	9.495
552	45	5.149	6.648	5.800	5.779
553	38	5.347	5.990	5.714	5.721
559	32	5.011	6.345	6.244	6.216
561	37	4.918	6.324	5.015	4.990
563	37	4.101	4.833	4.838	4.832
566	25	4.223	4.802	4.732	4.731
570	53	4.903	6.448	5.452	5.453
601	65	6.176	7.412	7.383	7.368
604	41	3.562	3.672	3.590	3.595
652	26	3.559	3.812	3.738	3.737
767	33	5.435	6.723	6.285	6.293
768	18	3.128	4.999	4.772	4.794
769	53	5.344	7.249	6.067	6.076
772	43	4.952	7.359	6.436	6.432
774	44	7.389	9.753	7.931	7.932
785	32	5.990	5.557	5.467	5.469
786	25	3.545	7.186	4.982	4.960
788	65	4.737	4.856	4.776	4.960
795	35	6.103	6.703	6.709	6.708
797	35	5.375	7.268	5.847	5.877
808	35	3.694	5.407	4.040	4.029
821	52	4.509	5.007	4.653	4.652
822	32	5.444	6.447	5.895	5.917
TOTAL	1946	4.890	6.152	5.430	5.430

Table 11

Correlations of Observed and Predicted Scores
For Cohort I (76-78)

School	N	Within LSQ	Pooled LSQ	Pooled LSQ with Adjusted α	Bayesian Equal Slope Model
51	63	0.930	0.921	0.920	0.920
101	24	0.919	0.913	0.914	0.914
102	47	0.918	0.912	0.911	0.911
106	35	0.942	0.935	0.935	0.935
205	26	0.780	0.743	0.748	0.748
206	52	0.883	0.879	0.880	0.880
207	25	0.918	0.896	0.897	0.897
209	44	0.924	0.917	0.916	0.916
220	24	0.871	0.821	0.827	0.826
225	12	0.857	0.761	0.766	0.766
226	72	0.911	0.894	0.898	0.898
235	74	0.913	0.892	0.896	0.896
302	30	0.913	0.869	0.866	0.866
304	23	0.948	0.921	0.922	0.922
307	11	0.947	0.928	0.927	0.927
308	53	0.935	0.923	0.922	0.922
310	59	0.928	0.927	0.927	0.927
312	38	0.854	0.817	0.314	0.314
313	63	0.904	0.889	0.887	0.887
409	13	0.853	0.579	0.577	0.577
414	32	0.941	0.939	0.910	0.910
415	24	0.931	0.910	0.910	0.910
416	19	0.942	0.828	0.837	0.836
422	32	0.865	0.855	0.853	0.853
504	46	0.937	0.936	0.936	0.936
505	63	0.907	0.900	0.902	0.901
509	41	0.519	0.480	0.478	0.479
552	45	0.918	0.894	0.896	0.896
553	38	0.888	0.878	0.880	0.880
559	32	0.899	0.863	0.863	0.863
561	37	0.859	0.848	0.849	0.849
563	37	0.948	0.936	0.935	0.935
566	25	0.956	0.936	0.936	0.936
570	53	0.896	0.869	0.869	0.869
601	65	0.842	0.777	0.779	0.779
604	41	0.925	0.900	0.904	0.903
652	26	0.949	0.945	0.946	0.946
767	33	0.805	0.795	0.795	0.795
768	18	0.979	0.969	0.969	0.969
769	53	0.909	0.885	0.883	0.883
772	43	0.861	0.811	0.813	0.813
774	44	0.852	0.840	0.839	0.839
785	32	0.841	0.821	0.818	0.818
786	25	0.942	0.888	0.885	0.885
788	65	0.922	0.918	0.918	0.918
795	35	0.869	0.848	0.849	0.849
797	35	0.900	0.874	0.878	0.878
808	35	0.938	0.922	0.924	0.924
821	52	0.933	0.930	0.931	0.931
822	32	0.920	0.903	0.901	0.902

lationship but simply fits idiosyncrasies found in the sample.

RESULTS OF THE CROSS-VALIDATION STUDIES

To gain more insight into the predictive efficiency of the regression equations it was decided to undertake a cross-validation study. A double cross-validation analysis was performed. For the first cross-validation, regression coefficients obtained from Cohort I were applied to Cohort II. For the second cross-validation this process was reversed with estimates of the regression coefficients based on Cohort II being applied to Cohort I. The regression coefficients obtained from Cohort II and used in the second cross-validation are found in Table 12.

Typically the findings in cross-validation studies show an increase in the error of prediction, on the average, when regression equations from one sample are used in a second sample. Similarly, the correlation between predicted values and observed values of the criterion are somewhat lower, on the average, than those obtained from the first sample. Again this is due to the impreciseness of the regression equations estimating the true relationship.

It was hypothesized that at least in theory the Bayesian simultaneous estimates would provide better estimates than

Table 12

Regression Estimates for Cohort II (77-79)

Schools	Within LSQ α	β_1	β_2	β_3	β_4	β_5	Pooled LSQ with Adjusted α α	Bayesian Equal Slope Model α
51	0.148	0.109	0.231	0.180	0.270	0.163	3.290	3.377
101	-1.508	-0.015	0.425	-0.076	0.352	0.264	0.834	1.190
102	15.132	0.169	0.311	-0.002	0.096	0.146	1.017	1.274
106	1.336	0.238	0.207	0.277	0.057	0.141	3.223	3.318
205	-10.690	0.329	0.060	0.220	0.242	0.213	0.495	0.803
206	5.397	-0.003	0.205	0.411	0.138	0.104	4.736	4.774
207	1.487	0.208	0.211	0.291	-0.407	0.671	5.955	5.944
209	-10.811	0.364	0.097	0.458	-0.057	0.236	2.085	2.256
220	15.484	0.009	0.171	-0.017	0.279	0.398	11.816	11.576
225	-8.995	-0.065	0.148	0.257	0.251	0.384	-4.459	-3.703
226	17.192	0.114	0.114	0.106	0.352	0.065	8.833	8.741
235	1.399	0.305	-0.010	0.305	0.131	0.221	5.885	5.891
302	5.870	-0.232	0.360	0.171	0.389	0.190	0.270	0.542
304	10.817	0.163	0.100	0.031	0.229	0.301	8.668	8.299
307	-8.407	-0.234	0.408	0.387	0.326	0.216	2.132	2.477
308	20.140	0.242	0.099	0.148	0.011	0.263	12.442	12.200
310	7.298	0.266	-0.015	0.047	0.260	0.305	4.882	4.925
312	10.694	0.091	0.042	0.261	0.398	0.024	5.675	5.681
313	17.639	0.114	0.210	0.187	0.113	0.184	12.496	12.167
409	-25.053	0.611	-0.679	1.000	0.116	0.389	9.556	9.004
414	17.191	0.049	0.169	0.166	0.313	0.092	9.015	8.863
415	39.002	-0.163	0.061	0.081	0.391	0.149	12.013	11.609
416	21.955	0.309	0.440	-0.312	0.059	0.159	5.955	5.860
422	4.344	0.149	0.136	0.179	0.377	0.085	6.938	6.907
504	-23.400	0.071	0.129	0.674	0.084	0.242	1.250	1.473
505	-5.827	0.224	0.118	0.275	0.230	0.160	2.472	2.591
509	-14.927	-0.079	0.157	0.569	0.316	0.191	5.654	5.635
552	4.775	0.070	0.143	0.138	0.299	0.193	1.790	1.954
553	5.146	0.467	-0.220	0.450	-0.082	0.316	6.528	6.482
559	10.932	0.268	0.113	-0.129	0.473	0.152	7.553	7.449
561	-1.942	0.033	0.302	0.369	0.115	0.143	2.595	2.703
563	-9.932	0.057	0.209	0.324	0.256	0.202	0.563	0.781
566	4.354	-0.004	0.267	0.014	0.179	0.467	5.350	5.340
570	5.166	0.137	0.168	0.127	0.243	0.187	2.689	2.804
601	25.734	-0.149	0.330	0.203	0.052	0.190	3.458	3.541
604	1.749	0.112	0.152	0.175	0.256	0.254	5.633	5.639
652	2.832	0.296	0.033	0.351	0.090	0.162	5.774	5.759
767	-7.854	0.136	0.272	0.375	-0.106	0.374	1.564	1.773
768	-17.695	0.227	0.189	0.530	-0.128	0.353	1.793	2.083
769	10.753	0.246	0.136	-0.003	0.171	0.223	3.428	3.540
772	13.373	0.235	0.009	0.201	0.209	0.154	8.512	8.425
774	-11.372	0.289	0.078	0.431	0.258	0.107	4.104	4.173
785	7.517	0.094	0.195	0.243	0.171	0.106	2.629	2.775
786	30.550	-0.052	0.124	0.142	0.268	0.155	15.309	14.627
788	5.865	0.039	-0.012	0.328	0.202	0.285	4.506	4.562
795	-8.269	0.189	-0.048	0.455	0.052	0.383	2.796	2.958
797	0.204	-0.069	0.403	0.405	0.099	0.132	4.830	4.877
808	12.288	0.299	0.072	0.007	0.076	0.356	5.714	5.702
821	2.884	-0.039	0.318	0.240	0.127	0.290	5.134	5.163
822	0.618	0.207	0.107	0.248	0.252	0.117	3.372	3.482

Pooled LSQ	α=5.293	β_1=0.128	β_2=0.177	β_3=0.211	β_4=0.157	β_5=0.212
Pooled LSQ with Adjusted α		β_1=0.125	β_2=0.150	β_3=0.233	β_4=0.164	β_5=0.218
Bayesian Equal Slope Model		β_1=0.125	β_2=0.152	β_3=0.231	β_4=0.163	β_5=0.218

would the within least squares regression equations. The Bayesian approach should moderate extreme values and thus improve the estimates of the regression equations by discounting idiosyncrasies in the initial sample.

The results of these cross-validation studies generally support initial expectations. Tables 13, 14, and 15 show comparisons of mean square errors, absolute errors, and correlations between observed and predicted scores within each school and for all subjects from the first cross-validation analysis. Errors of prediction, on the average, increased for all regression models except the pooled least squares regression. As expected, the Bayesian m-group equal slope regression model proved more effective than the within least squares regression model. However, the pooled least squares regression with adjusted alpha was almost as effective and the pooled least squares regression was slightly better. The results of this cross-validation analysis confirm the results of Lissitz and Schoenfeldt (1974). They reported that although some improvement in prediction was obtained by Bayesian m-group regression analysis over the within least squares regression, no improvement was offered when compared to the total group (pooled) least squares procedure.

The second cross-validation analysis produced slightly different results. Comparisons of the mean squared errors,

Table 13

Mean Square Errors for Crossvalidation I

School	N	Within LSQ	Pooled LSQ	Pooled LSQ with Adjusted α	Bayesian Equal Slope Model
51	60	62.227	63.387	62.303	62.406
101	23	62.661	90.867	70.460	71.466
102	33	89.114	61.268	55.111	54.964
106	55	71.565	65.609	68.128	67.726
205	31	112.943	75.002	58.177	58.998
206	39	42.047	35.461	34.374	33.922
207	38	103.905	96.407	99.749	99.394
209	40	73.182	62.981	95.005	93.024
220	44	106.738	64.312	128.907	123.043
225	23	96.842	167.973	63.426	72.538
226	56	42.683	33.367	31.101	31.167
235	57	41.031	35.472	34.664	34.702
302	35	83.646	100.217	65.671	65.737
304	15	73.779	48.819	143.349	131.023
307	16	48.187	69.092	47.639	49.331
308	49	39.785	71.851	37.550	37.918
310	66	68.088	58.553	58.243	58.049
312	42	78.030	45.011	85.735	81.635
313	37	19.053	50.850	17.915	18.325
409	13	308.010	58.203	146.607	130.013
414	38	25.466	26.154	18.165	18.474
415	28	79.283	98.793	65.827	64.419
416	12	37.024	22.048	30.690	29.337
422	53	65.206	30.608	50.375	47.380
504	37	31.247	61.113	37.476	37.290
505	59	75.110	53.591	73.256	72.282
509	18	81.393	36.199	34.884	34.966
552	43	36.076	50.156	25.981	26.175
553	37	64.204	66.732	68.979	68.787
559	33	68.912	59.504	58.408	58.022
561	58	61.381	76.121	61.607	61.269
563	45	86.181	65.882	46.767	47.342
566	15	24.659	38.593	37.427	37.446
570	62	52.341	50.544	37.794	37.787
601	68	78.421	86.550	84.009	83.648
604	36	22.776	22.569	21.560	21.604
652	27	22.767	21.007	19.989	19.999
767	35	114.018	68.798	97.665	95.351
768	22	70.216	68.548	61.519	61.935
769	38	59.979	43.583	36.613	35.997
772	52	55.798	44.658	52.041	50.389
774	45	161.344	73.932	161.988	155.463
785	38	81.671	47.687	58.443	57.381
786	25	80.535	114.862	46.319	48.731
788	47	31.794	35.273	31.383	31.391
795	30	45.762	44.287	41.915	41.816
797	50	120.840	59.335	92.429	89.436
808	28	83.723	44.476	73.572	70.544
821	45	52.643	40.034	50.127	49.444
822	37	59.456	39.809	57.940	56.229
TOTAL	1933	68.409	58.286	60.030	59.137

Table 14

Mean Absolute Errors for Crossvalidation I

School	N	Within LSQ	Pooled LSQ	Pooled LSQ with Adjusted α	Bayesian Equal Slope Model
51	60	6.017	6.638	6.299	6.305
101	23	5.621	6.889	5.623	5.671
102	33	7.715	6.142	5.857	5.847
106	55	6.554	6.145	6.281	6.262
205	31	9.104	7.397	6.345	6.404
206	39	5.010	4.504	4.590	4.557
207	38	6.518	6.305	7.269	7.209
209	40	6.895	5.743	7.679	7.570
220	44	8.975	6.758	9.990	9.713
225	23	8.013	11.111	6.048	6.572
226	56	5.078	4.519	4.283	4.292
235	57	5.001	4.799	4.739	4.740
302	35	7.346	7.560	6.031	6.044
304	15	6.778	5.779	9.905	9.349
307	15	5.709	7.641	5.953	6.146
308	49	4.869	6.688	4.570	4.597
310	66	6.315	5.608	5.999	5.980
312	42	7.364	5.473	7.630	7.414
313	37	3.619	6.298	3.373	3.409
409	13	15.254	6.006	10.385	9.640
414	38	4.213	4.125	3.524	3.554
415	28	7.398	7.917	6.970	6.960
416	12	4.850	4.348	4.971	4.888
422	53	5.900	4.145	5.432	5.179
504	37	4.168	5.744	4.682	4.648
505	59	6.967	5.690	6.865	6.808
509	18	7.691	5.237	5.173	5.178
552	43	4.942	5.792	4.167	4.156
553	37	6.262	6.422	6.460	6.461
559	33	6.381	5.841	5.558	5.549
561	68	6.142	6.583	6.010	5.985
563	45	7.795	6.283	5.148	5.181
566	15	4.260	5.061	4.777	4.779
570	62	6.009	5.595	4.836	4.832
601	68	6.747	7.339	7.055	7.044
604	36	3.938	3.780	3.671	3.674
652	27	3.882	3.951	3.818	3.817
767	35	8.687	6.760	8.254	8.142
768	22	6.725	7.198	6.732	6.761
769	38	6.368	5.089	4.868	4.823
772	52	5.795	5.195	5.887	5.741
774	45	11.409	7.434	11.637	11.422
785	38	7.140	5.464	6.183	6.117
786	25	6.460	8.846	4.864	5.072
788	47	4.757	4.737	4.564	4.562
795	30	5.222	5.340	5.177	5.178
797	50	8.044	5.445	6.907	6.741
808	28	6.738	5.421	6.549	6.399
821	45	6.044	5.245	5.931	5.894
822	37	6.089	4.853	6.311	6.193
TOTAL	1933	6.393	5.902	5.987	5.943

Table 15

Correlations of Observed and Predicted Scores
For Crossvalidation I

School	N	Within LSQ	Pooled LSQ	Pooled LSQ with Adjusted α	Bayesian Equal Slope Model
51	60	0.884	0.894	0.893	0.843
101	23	0.899	0.897	0.896	0.896
102	33	0.852	0.863	0.858	0.859
106	55	0.909	0.920	0.918	0.918
205	31	0.898	0.914	0.916	0.916
206	39	0.923	0.931	0.930	0.930
207	38	0.823	0.833	0.829	0.829
209	40	0.916	0.914	0.912	0.912
220	44	0.817	0.810	0.812	0.812
225	23	0.659	0.825	0.829	0.828
226	56	0.899	0.910	0.912	0.912
235	57	0.862	0.884	0.887	0.887
302	35	0.855	0.882	0.880	0.880
304	15	0.855	0.886	0.886	0.886
307	16	0.947	0.954	0.953	0.953
308	49	0.871	0.865	0.865	0.865
310	66	0.882	0.887	0.889	0.889
312	42	0.817	0.869	0.872	0.872
313	37	0.917	0.928	0.927	0.927
404	13	0.850	0.913	0.912	0.912
414	38	0.924	0.926	0.927	0.927
415	28	0.536	0.615	0.621	0.621
416	12	0.879	0.930	0.929	0.929
422	53	0.887	0.906	0.908	0.908
504	37	0.913	0.907	0.907	0.907
505	59	0.884	0.887	0.888	0.888
509	18	0.731	0.892	0.894	0.894
552	43	0.933	0.949	0.950	0.950
553	37	0.890	0.884	0.886	0.886
559	33	0.850	0.868	0.871	0.870
561	58	0.877	0.881	0.880	0.880
563	45	0.937	0.949	0.949	0.949
566	15	0.967	0.939	0.937	0.937
570	62	0.885	0.899	0.898	0.898
601	68	0.714	0.713	0.708	0.709
604	36	0.915	0.925	0.926	0.926
752	27	0.950	0.955	0.957	0.957
767	35	0.915	0.922	0.921	0.921
768	22	0.925	0.936	0.934	0.934
769	38	0.841	0.884	0.883	0.883
772	52	0.904	0.922	0.922	0.922
774	45	0.917	0.933	0.933	0.933
785	38	0.903	0.930	0.930	0.930
786	25	0.723	0.846	0.848	0.847
788	47	0.935	0.936	0.938	0.938
795	30	0.946	0.953	0.954	0.954
797	50	0.817	0.887	0.885	0.885
808	28	0.885	0.901	0.900	0.900
821	45	0.908	0.913	0.911	0.911
822	37	0.881	0.900	0.900	0.900

absolute errors, and correlations between observed and predicted scores within a school and for all subjects in the second cross-validation analysis are reported in Tables 16, 17, and 18, respectively. The results obtained in this cross-validation analysis indicated that the Bayesian model and the pooled least squares regression model with adjusted alpha were clearly better than the within least squares regression approach in terms of predictive efficiency but provided only a slightly better prediction system than the pooled least squares regression. Unlike the results of Novick and Jackson (1974) and Shigemasu (1976) where the pooled least squares regression showed about half of the improvement over the within least squares regression attained by the Bayesian models, the predictive efficiency of the pooled least squares regression method in this analysis more closely resembled the Bayesian approach.

Table 19 provides a comparison of the mean square errors when the mean square error for the Bayesian m-group equal slope regression analysis is set to be 100. It is obvious that the predictive accuracy of the within-school least squares regression method is somewhat inferior to the remaining three methods. However, it appears that the predictive accuracy of the Bayesian m-group equal slope model shows no appreciable advantage over the pooled least squares

Table 16

Mean Square Errors for Crossvalidation II

School	N	Within LSQ	Pooled LSQ	Pooled LSQ with Adjusted α	Bayesian Equal Slope Model
51	63	48.768	46.464	50.242	50.217
101	24	69.965	38.419	52.873	50.648
102	47	72.220	51.246	69.397	67.882
106	35	73.779	50.947	61.689	61.371
205	26	95.356	58.876	77.926	75.862
206	52	72.994	67.531	65.658	65.653
207	25	112.085	74.644	74.693	74.634
209	44	96.290	49.569	79.730	78.407
220	24	136.448	35.855	128.432	121.893
225	12	120.497	132.413	121.013	116.031
226	72	38.387	39.199	35.266	34.844
235	74	38.081	38.552	36.347	36.418
302	30	66.690	79.522	59.851	59.745
304	23	85.218	68.243	133.555	124.467
307	11	65.422	44.611	44.684	43.991
308	53	31.904	70.467	28.105	27.608
310	59	46.889	29.949	30.393	30.255
312	38	130.191	122.162	113.635	114.210
313	63	43.222	82.365	47.211	46.132
409	13	219.007	124.561	196.095	183.721
414	32	36.856	31.712	22.407	22.100
415	24	86.214	106.006	33.741	35.411
416	19	136.660	77.928	67.471	68.533
422	32	59.234	69.145	47.922	48.790
504	46	32.711	43.986	26.572	26.794
505	63	69.685	49.719	71.148	70.693
509	41	250.336	186.824	188.598	188.403
552	45	59.207	68.876	60.318	60.342
553	38	82.992	46.277	47.239	47.074
559	32	60.129	85.533	69.732	70.213
561	37	50.593	61.046	49.487	49.636
563	37	61.624	30.529	47.159	45.961
566	25	30.344	34.804	34.910	34.888
570	53	46.609	54.037	49.420	49.397
601	65	79.888	92.419	87.798	87.844
604	41	24.026	24.043	23.471	23.490
652	26	28.811	21.484	20.024	20.104
767	33	170.828	95.261	132.733	130.644
768	18	122.158	28.281	44.259	42.494
769	53	65.257	62.672	54.591	54.733
772	43	57.059	115.324	76.888	77.934
774	44	215.777	171.586	185.556	185.461
785	32	97.964	81.056	99.465	98.498
786	25	60.869	101.716	45.536	42.241
788	65	50.910	39.160	39.914	39.949
795	35	123.861	72.459	81.888	81.213
797	35	130.259	97.566	94.651	95.200
808	35	57.758	62.202	54.302	55.534
821	52	50.260	47.173	45.186	45.467
822	32	98.747	72.498	86.786	86.194
TOTAL	1946	75.974	66.319	65.455	64.904

Table 17

Mean Absolute Errors for Crossvalidation II

School	N	Within LSQ	Pooled LSQ	Pooled LSQ with Adjusted α	Bayesian Equal Slope Model
51	63	5.775	5.630	5.871	5.871
101	24	6.843	5.123	6.022	5.878
102	47	7.085	5.978	6.946	6.362
106	35	6.715	5.912	6.394	6.389
205	26	7.759	6.556	7.490	7.393
206	52	7.060	6.478	6.377	6.377
207	25	8.100	6.435	6.512	6.505
209	44	8.066	5.880	7.677	7.602
220	24	10.466	4.246	10.047	9.715
225	12	9.350	9.843	8.874	8.875
226	72	4.544	5.040	4.322	4.295
235	74	4.926	4.828	4.772	4.775
302	30	6.299	7.874	6.728	6.759
304	23	7.600	6.315	10.092	9.628
307	11	7.394	5.028	5.271	5.203
308	53	4.046	7.346	3.916	3.886
310	59	5.051	4.291	4.323	4.318
312	38	9.353	9.039	8.512	8.552
313	63	5.196	7.466	5.548	5.457
409	13	12.791	8.337	11.065	10.542
414	32	4.491	4.846	3.653	3.673
415	24	8.043	9.007	4.561	4.701
416	19	7.978	5.175	4.675	4.712
422	32	6.140	7.240	5.731	5.793
504	46	4.258	5.493	4.446	4.453
505	63	6.802	5.683	6.999	6.974
509	41	11.666	9.556	9.531	9.532
552	45	6.021	6.163	6.024	6.018
553	38	6.981	5.819	5.867	5.856
559	32	5.989	7.066	6.280	6.285
561	37	5.718	5.729	5.142	5.153
563	37	6.369	4.556	5.797	5.722
566	25	4.296	4.762	4.740	4.745
570	53	5.277	5.865	5.449	5.451
601	65	7.068	7.570	7.520	7.521
604	41	3.661	3.757	3.605	3.614
652	26	4.176	3.812	3.707	3.707
767	33	10.334	7.579	9.774	9.674
768	18	9.481	4.422	5.367	5.239
769	53	6.591	6.510	6.173	6.177
772	43	6.109	8.512	7.090	7.126
774	44	11.515	10.672	11.110	11.110
785	32	6.974	5.665	6.545	6.504
786	25	6.100	8.322	5.715	5.407
788	65	5.726	4.966	5.050	5.055
795	35	8.042	6.721	6.921	6.903
797	35	9.471	8.388	8.264	8.292
808	35	6.252	6.647	6.109	6.165
821	52	5.842	5.722	5.580	5.602
822	32	8.124	6.995	7.904	7.871
TOTAL	1946	6.642	6.333	6.248	6.221

Table 18

Correlations of Observed and Predicted Scores
For Crossvalidation II

School	N	Within LSQ	Pooled LSQ	Pooled LSQ with Adjusted α	Bayesian Equal Slope Model
51	63	0.920	0.918	0.918	0.918
101	24	0.884	0.914	0.914	0.914
102	47	0.882	0.910	0.912	0.912
106	35	0.929	0.933	0.933	0.933
205	26	0.730	0.733	0.737	0.737
206	52	0.859	0.880	0.881	0.881
207	25	0.839	0.894	0.897	0.897
209	44	0.911	0.917	0.919	0.919
220	24	0.835	0.812	0.814	0.814
225	12	0.753	0.761	0.765	0.765
226	72	0.890	0.890	0.895	0.895
235	74	0.888	0.893	0.893	0.893
302	30	0.842	0.867	0.865	0.865
304	23	0.916	0.922	0.925	0.925
307	11	0.939	0.929	0.927	0.927
308	53	0.927	0.925	0.926	0.926
310	59	0.899	0.926	0.926	0.926
312	38	0.758	0.825	0.821	0.821
313	63	0.896	0.892	0.889	0.889
409	13	0.800	0.567	0.586	0.585
414	32	0.925	0.939	0.938	0.938
415	24	0.756	0.911	0.912	0.912
416	19	0.671	0.824	0.832	0.831
422	32	0.829	0.855	0.856	0.856
504	46	0.928	0.935	0.935	0.935
505	63	0.901	0.899	0.902	0.901
509	41	0.454	0.480	0.475	0.475
552	45	0.894	0.892	0.895	0.895
553	38	0.780	0.872	0.876	0.876
559	32	0.877	0.867	0.865	0.865
561	37	0.846	0.843	0.846	0.846
563	37	0.930	0.938	0.939	0.939
566	25	0.952	0.938	0.937	0.937
570	53	0.874	0.868	0.865	0.865
601	65	0.787	0.769	0.771	0.771
604	41	0.899	0.900	0.900	0.900
752	26	0.924	0.945	0.945	0.945
767	33	0.773	0.793	0.793	0.793
768	18	0.953	0.970	0.970	0.970
769	53	0.858	0.884	0.883	0.883
772	43	0.829	0.807	0.812	0.812
774	44	0.831	0.842	0.843	0.843
785	32	0.820	0.826	0.823	0.823
786	25	0.848	0.892	0.887	0.887
788	65	0.903	0.916	0.917	0.917
795	35	0.834	0.844	0.846	0.846
797	35	0.843	0.878	0.874	0.873
808	35	0.867	0.920	0.922	0.922
821	52	0.912	0.930	0.931	0.931
822	32	0.900	0.904	0.904	0.904

regression with adjusted alpha or the pooled least squares regression model with no adjustments.

COMPARISON OF THE SCHOOL EFFECTIVENESS INDICES

Nine school effectiveness indices were derived from the six models used in this study. Three indices were computed using the within-school least squares regression model. These indices are regression estimates of the mean posttest composite scores at three reference points, the overall mean of individual pretest subtest scores and points one standard deviation above and below the mean. The index produced from the pooled least squares regression model represents the average of the residuals for the individuals within each school. School effectiveness indices for the pooled least squares regression with adjusted alphas model as well as the Bayesian m-group equal slope model represent regression estimates of the mean posttest composite score at reference points equivalent to the overall pretest subtest means. School residuals obtained by subtracting the observed mean posttest score from the predicted mean posttest score served as one index using the school means least squares regression model. The other effectiveness index obtained from this model was the SEI classification developed by Dyer, Linn and Patton (1967). Finally, the computation of mean difference

Table 19

A Comparison of Mean Squared Error When MSE=100
For Bayesian M-Group Equal Slope Analysis

Regression Analysis	Crossvalidation I	Crossvalidation II
Within LSQ	115.68	117.06
Pooled LSQ	98.56	102.18
Pooled LSQ With Adjusted α	101.51	100.85
Bayesian M-Group Equal Slope Model	100.00	100.00

scores for each school served as an additional school effectiveness measure.

These indices are reported in Table 20 for each school in Cohort I. While these school effectiveness indices are not comparable in the absolute sense, the relative positions of the schools in relation to other schools in the study may be compared using these indices. These indices were compared directly using correlations. In addition, indirect comparisions were made by comparing the correlations of these indices with other school and community variables.

Direct Comparisions of Indices

The intercorrelations among the school effectiveness indices obtained from Cohort I are reported in Table 21. The six methods of estimating school effectiveness indices yielded indices that were highly correlated. Only one correlation, the correlation of the school effectiveness index for low-scoring students with the index for high-scoring students (r=.51) was below .75.

It should be noted that the school effectiveness indices produced using the Bayesian m-group equal slope model, the pooled least squares regression model and the pooled least squares model with adjusted alphas correlated nearly perfectly with each other (r=.99). This is not surprising con-

Table 20

School Effectiveness Indices For Cohort I (76-78)

School	Within LSQ (L)	Within LSQ (M)	Within LSQ (H)	Pooled LSQ With Adjusted α	Bayesian Equal Slope Model	Individual Residuals	School Residuals	Dyer's SEIs	Mean Differences
51	41.019	58.866	76.713	59.579	59.594	-1.205	-0.516	3	-2.040
101	42.802	58.066	73.330	58.723	58.842	-2.197	-2.122	2	-1.189
102	39.877	58.922	77.966	59.731	59.750	-1.061	-0.859	3	-2.048
106	45.042	61.166	77.290	61.195	61.134	0.252	0.507	3	2.031
205	45.274	57.907	70.540	59.124	59.208	-1.721	-1.721	2	-1.448
206	40.364	57.630	74.897	57.074	57.169	-3.946	-5.406	1	-2.276
207	41.472	59.746	73.021	58.782	58.898	-2.016	-1.451	2	-1.529
209	45.647	63.300	80.954	63.416	63.280	2.690	3.717	5	1.721
220	42.055	57.876	73.697	56.466	56.783	-4.052	-2.626	2	-6.254
225	41.859	56.935	72.011	54.096	55.020	-6.520	-4.976	1	-8.591
226	43.402	60.865	78.327	61.517	61.482	0.842	1.947	4	-0.716
235	43.399	60.230	77.061	60.726	60.712	0.071	1.151	3	-1.341
302	43.978	58.377	72.775	54.869	55.213	-5.794	-5.845	1	-6.151
304	40.174	55.304	70.433	52.583	53.311	-7.945	-4.686	1	-10.154
307	42.145	57.191	72.237	57.366	57.845	-3.421	-2.111	2	-2.326
308	48.106	65.411	82.715	65.924	65.708	5.035	3.596	5	4.600
310	38.557	56.859	75.162	57.723	57.303	-2.993	-3.727	1	-4.521
312	48.781	65.016	81.251	66.600	66.276	5.704	4.451	5	6.163
313	49.395	65.601	81.806	65.451	65.288	4.705	5.075	5	3.530
409	32.244	52.313	72.383	53.934	54.610	-6.775	-7.932	1	-8.713
414	41.969	61.318	80.667	62.725	62.589	2.081	2.357	4	0.625
415	50.579	67.311	84.044	68.080	67.497	7.415	7.745	5	5.977
416	46.770	61.685	76.600	63.071	62.831	2.450	3.469	4	0.828
422	54.635	68.135	81.635	66.071	65.739	5.373	4.509	5	3.202
504	35.254	53.354	71.455	55.072	55.296	-5.539	-4.281	1	-8.087
505	43.744	61.942	80.140	62.744	62.669	2.124	3.220	4	0.410
509	49.709	60.589	71.470	59.342	59.867	-0.773	0.943	3	-1.696
552	41.779	56.117	70.454	56.571	56.721	-4.291	-4.434	1	-4.422
553	43.327	61.021	78.714	58.714	58.777	-2.276	-1.182	2	-0.074
559	49.407	64.121	78.835	63.474	63.290	2.692	3.679	5	2.366
561	41.711	57.141	72.571	56.067	56.285	-4.633	-4.575	1	-5.459
563	39.950	59.825	79.701	58.923	58.989	-1.912	-2.088	2	-2.373
566	43.228	59.075	74.922	59.881	59.899	-1.214	-3.154	2	0.593
570	38.666	57.694	76.723	56.974	57.093	-3.773	-4.000	1	-5.777
601	43.534	58.811	74.087	57.212	57.317	-3.152	-0.086	3	-5.825
604	42.325	59.751	77.177	60.268	60.274	-0.404	-0.271	3	-2.191
652	44.529	60.263	75.998	60.585	60.559	-0.386	-1.025	3	0.909
767	50.388	63.061	75.739	63.189	63.024	2.388	2.170	4	1.931
768	39.794	60.759	81.724	59.878	59.932	-0.907	0.494	3	-0.924
769	38.140	55.343	72.546	55.615	55.777	-5.185	-4.669	1	-5.542
772	50.912	62.364	73.816	65.874	65.607	4.799	4.022	5	6.512
774	49.839	67.052	84.265	68.227	67.869	7.228	5.363	5	7.725
785	48.103	62.716	77.329	62.006	61.900	1.038	-1.183	2	0.730
786	53.812	68.810	83.808	67.304	66.793	6.520	6.551	5	6.200
788	42.925	59.663	76.401	59.947	59.942	-0.962	-1.500	2	-0.601
795	45.032	60.279	75.527	60.669	60.641	-0.247	-1.451	2	-0.226
797	44.557	63.402	82.247	65.102	64.851	4.450	6.399	5	3.299
808	50.686	66.065	81.445	65.509	65.220	4.597	3.178	4	4.678
821	46.840	62.810	78.779	62.814	62.718	2.102	2.842	4	0.897
822	46.049	62.854	79.659	62.979	62.811	2.003	0.292	3	2.702

Table 21

Intercorrelations Among School Effectiveness Indices
Computed by Six Methods for Cohort I (76-78)

	School Effectiveness Index								
School Effectiveness Index	1	2	3	4	5	6	7	8	9
1. Within LSQ (L)	1.000	0.889	0.509	0.817	0.817	0.817	0.789	0.761	0.804
2. Within LSQ (M)		1.000	0.846	0.942	0.942	0.944	0.908	0.879	0.905
3. Within LSQ (H)			1.000	0.822	0.821	0.824	0.790	0.768	0.767
4. Pooled LSQ with Adjusted α				1.000	0.999	0.999	0.941	0.924	0.966
5. Bayesian M-Group Equal Slope					1.000	0.999	0.941	0.927	0.965
6. Individual Residuals						1.000	0.950	0.932	0.958
7. School Residuals							1.000	0.966	0.869
8. Dyer's SEIs								1.000	0.853
9. Mean Differences									1.000
Mean	44.28	60.62	76.96	60.61	60.62	-0.17	0.00	2.92	-0.71
S.D.	4.57	3.70	3.93	3.96	3.70	3.92	3.73	1.47	4.29

sidering the results of predictive efficiency of these models obtained in the previous section.

In addition, these three indices correlated higher than the two indices based on school residuals or mean difference scores with the three within least squares regression indices. Not surprisingly, the correlations of these three indices with the index for the middle-scoring students (r=.94) were considerably higher than the correlations with indices for low- and high-scoring students (r=.82 for both).

It was also found that the school effectiveness indices produced using low, middle, and high scores produced substantially different results. Although the correlations of the effectiveness index for middle-scoring students with those for low- and high-scoring students were relatively high (r=.89 and r=.85, respectively), the correlation of indices for the low-scoring students with high-scoring students was only .51. These results confirm those given by Marco (1974) were data suggested that "more than one school effectiveness index may be needed to accurately describe the effectiveness of a school for a given group of students (Marco, 1974, p. 230)."

Indirect Comparisions of Indices

Correlations of school effectiveness indices for Cohort I with twelve other school variables are reported in Table 22. A correlation of ±.275 is significantly different from zero at α=.05. Only two school variables have correlations that approach significance. These variables are the average pupil attendance and a property destruction variable.

The average pupil attendance represents the ratio of the average daily atendance to the average number belonging. It appears that schools with higher school effectiveness indices tend to have a higher attendance percentage with correlations ranging from .11 to .32. This relationship holds for all school effectiveness indices but is not as strong a relationship for the mean difference scores and the within least squares index for low-scoring students.

The property destruction variable is a monetary figure that indicates the total gross loss to schools including glass breakage, theft, miscellaneous destruction and fire costs for FY79. Those schools with less property destruction tended to have higher school effectiveness indices. This relationship is strongest for the mean difference scores (r=-.32) and weaker for those indices using within least squares regression and mean school residuals with correlations ranging from -.15 to -.23.

Table 22

Correlations of School Effectiveness Indices for Cohort I (76-78)
With Twelve Other Non-Achievement Variables

School Effectiveness Index	Non-Achievement Variables											
	1	2	3	4	5	6	7	8	9	10	11	12
1. Within LSQ (L)	0.116	-0.057	-0.034	0.004	0.099	0.029	0.002	-0.087	-0.041	0.060	-0.145	-0.190
2. Within LSQ (M)	0.243	-0.125	-0.031	-0.005	0.179	0.067	-0.021	-0.102	-0.030	0.083	-0.210	-0.172
3. Within LSQ (H)	0.321	-0.170	0.039	-0.014	0.222	0.091	-0.042	-0.090	-0.009	0.086	-0.227	-0.102
4. Pooled LSQ with Adjusted α	0.265	-0.177	-0.039	0.011	0.188	0.083	-0.055	-0.070	-0.040	0.050	-0.274	-0.150
5. Bayesian M-Group Equal Slope	0.264	-0.171	-0.047	0.010	0.178	0.071	-0.042	-0.070	-0.034	0.057	-0.271	-0.150
6. Individual Residuals	0.290	-0.190	-0.033	0.014	0.196	0.072	-0.058	-0.075	-0.046	0.045	-0.271	-0.154
7. School Residuals	0.285	-0.193	0.021	0.044	0.225	0.041	-0.024	-0.088	-0.060	-0.027	-0.176	-0.159
8. Dyer's SEIs	0.289	-0.173	0.042	0.048	0.181	0.038	-0.077	-0.085	-0.107	-0.056	-0.179	-0.134
9. Mean Differences	0.140	-0.080	-0.050	-0.051	0.099	0.121	-0.029	-0.031	-0.030	0.078	-0.316	-0.118
Mean	95.80	21.41	46.20	25.43	17.33	388.16	21.72	12.94	11.38	18.41	1354.50	2.98
S.D.	0.85	9.46	28.05	1.89	1.94	125.27	12.03	7.23	7.72	7.07	2601.97	3.39

Note. The key to the numbers (1) through (12) are:

1. Average Pupil Attendance
2. Mobility Rate
3. Pupils Transported
4. Average Class Size
5. Student/Professional Staff Ratio
6. Total Student Enrollment
7. Percent of Minority Students
8. Staff Experience
9. Percent of Minority Staff
10. Sex of Staff Members
11. Property Destruction
12. Special Education Availability

It appears that school variables selected in this study provide very little additional information for interpreting school effectiveness indices based on academic achievement. Where some relationship did exist, the correlational patterns with the school effectiveness indices did not differ substantially.

STABILITY OF SCHOOL EFFECTIVENESS INDICES ACROSS SAMPLES

Students from Cohort I were divided into random halves within each school. Regression analyses were performed independently on each random subsample. The regression estimates for Sample I and Sample II are presented in Table 23 and Table 24 respectively. Regression estimates for the within-school least squares regression were not computed for five schools because of the limited number of students in these schools for each sample.

Nine school effectiveness indices were computed for all but these five schools using each subsample independently. For these five schools only those six school effectiveness indices that did not utilize the within least squares regression model were computed. The school effectiveness indices for Sample I and Sample II are given in Table 25 and Table 26, respectively.

Table 23

Regression Estimates for Sample I from Cohort I (76-78)

Schools	Within LSQ α	β_1	β_2	β_3	β_4	β_5	Pooled LSQ with Adjusted α α	Bayesian Equal Slope Model α
51	-4.554	0.013	0.281	0.513	0.283	-0.057	5.993	6.289
101	2.694	0.060	-0.129	-0.183	0.632	0.558	7.570	7.701
102	9.903	-0.022	0.337	0.034	0.083	0.409	7.272	7.469
106	5.389	0.134	0.134	0.252	0.304	0.080	7.294	7.474
205	-19.126	0.417	-0.193	0.152	0.608	0.305	5.043	5.662
206	5.053	-0.133	0.474	0.241	-0.104	0.353	3.954	4.432
207	4.191	-0.463	0.224	0.702	0.143	0.338	5.573	6.115
209	-5.079	0.432	0.000	0.283	-0.131	0.564	11.490	11.216
220	5.679	0.260	0.062	0.049	0.407	0.091	5.626	6.187
225	------	-----	-----	-----	-----	-----	2.990	4.611
226	5.121	0.002	0.006	0.291	0.519	0.090	8.907	9.001
235	8.063	0.099	0.158	-0.066	0.634	0.026	7.760	7.929
302	6.535	-0.140	0.497	0.447	0.322	-0.253	2.830	3.780
304	-5.319	0.208	0.151	0.660	-0.534	0.439	-0.374	1.405
307	------	-----	-----	-----	-----	-----	3.968	5.433
308	7.689	0.123	0.304	0.303	-0.084	0.306	13.562	13.156
310	-0.881	0.129	0.114	0.389	0.081	0.228	4.967	5.385
312	21.333	0.368	0.551	-0.410	0.190	0.019	13.895	13.233
313	2.162	0.298	0.098	0.156	0.380	0.065	10.701	10.616
409	------	-----	-----	-----	-----	-----	0.766	2.901
414	-10.320	0.227	0.277	0.312	0.312	0.045	10.549	10.296
415	18.584	0.426	0.176	-0.036	-0.074	0.308	14.672	13.496
416	------	-----	-----	-----	-----	-----	12.154	11.225
422	34.346	0.003	0.168	0.226	-0.027	0.202	12.307	11.790
504	-3.598	-0.211	0.248	0.331	0.230	0.324	2.017	2.809
505	-9.026	0.071	0.186	0.501	-0.025	0.401	9.122	9.191
509	26.995	0.159	0.153	0.016	-0.076	0.279	6.079	6.437
552	7.617	-0.107	0.048	0.176	0.053	0.585	2.425	3.136
553	-2.016	-0.016	0.079	0.592	0.044	0.309	5.771	6.129
559	7.878	0.711	-0.009	-0.100	0.067	0.211	8.539	8.559
561	9.580	0.080	0.021	0.449	0.044	0.186	3.435	4.180
563	-0.551	0.269	0.357	0.050	0.039	0.256	5.404	5.858
566	15.861	0.649	-0.074	-0.381	0.089	0.408	8.003	8.030
570	-4.161	0.657	-0.046	0.233	0.163	0.043	6.038	6.357
601	0.137	-0.303	0.304	0.301	0.526	0.093	2.357	2.984
604	2.833	0.274	0.091	0.089	0.467	0.020	6.885	7.149
652	1.566	-0.025	0.277	0.556	0.056	0.120	7.867	7.961
767	0.415	-0.175	0.133	0.650	0.750	-0.386	8.030	8.135
768	------	-----	-----	-----	-----	-----	7.701	7.832
769	-1.105	-0.167	0.297	0.597	-0.013	0.238	3.218	3.816
772	24.751	0.371	-0.233	0.337	0.005	0.138	12.174	11.787
774	11.940	0.164	0.500	0.301	-0.137	0.061	13.148	12.659
785	36.160	0.099	0.282	-0.527	0.479	0.194	11.471	11.028
786	30.721	0.380	0.611	-0.303	0.072	-0.127	13.759	12.829
788	7.066	0.209	0.208	0.311	0.008	0.103	6.170	6.430
795	-8.231	-0.068	0.459	0.415	0.177	0.099	5.969	6.332
797	-15.475	0.348	-0.431	0.737	0.421	0.168	11.480	11.113
808	16.593	0.016	0.017	0.008	0.047	0.717	12.607	12.044
821	13.514	0.142	0.074	0.040	0.348	0.206	9.736	9.696
822	-1.165	0.319	-0.057	0.742	-0.409	0.418	7.451	7.622

Pooled LSQ	α=8.596	β_1=0.100	β_2=0.173	β_3=0.245	β_4=0.154	β_5=0.178
Pooled LSQ with Adjusted α		β_1=0.120	β_2=0.151	β_3=0.246	β_4=0.178	β_5=0.171
Bayesian Equal Slope Model		β_1=0.118	β_2=0.154	β_3=0.245	β_4=0.175	β_5=0.173

Table 24

Regression Estimates for Sample II from Cohort I (76-78)

Schools	Within LSQ α	β_1	β_2	β_3	β_4	β_5	Pooled LSQ with Adjusted α α	Bayesian Equal Slope Model α
51	0.393	-0.125	0.145	0.516	0.197	0.224	7.074	7.238
101	11.161	0.297	0.157	0.045	0.087	0.154	3.816	4.361
102	-23.558	0.064	0.021	0.699	0.245	0.297	5.925	6.175
106	10.837	0.127	0.364	0.052	0.561	-0.254	9.145	9.128
205	25.291	0.041	0.013	0.501	0.110	-0.145	6.761	6.984
206	2.449	0.174	-0.082	0.174	0.439	0.200	4.235	4.516
207	-22.637	0.119	-0.238	0.713	0.313	0.417	5.723	6.061
209	13.048	0.134	0.322	0.262	0.036	0.073	9.335	9.363
220	14.937	0.003	0.137	-0.132	0.228	0.421	1.068	1.968
225	------	-----	-----	-----	-----	-----	-1.349	0.632
226	0.021	0.133	0.003	0.444	0.130	0.279	7.954	8.091
235	3.679	0.218	0.090	0.097	0.179	0.338	7.665	7.812
302	8.864	-0.127	0.044	0.979	-0.482	0.389	0.563	1.340
304	13.285	0.508	-0.058	-0.090	0.105	0.261	-0.807	0.323
307	------	-----	-----	-----	-----	-----	4.495	5.252
308	7.538	0.138	-0.066	0.332	0.094	0.423	12.049	11.942
310	-3.521	0.021	0.303	0.181	0.256	0.231	4.329	4.630
312	22.436	-0.412	0.229	0.499	0.065	0.351	11.806	12.912
313	25.865	0.246	0.380	-0.094	-0.015	0.184	10.916	13.933
409	------	-----	-----	-----	-----	-----	0.961	2.425
414	3.343	-0.131	0.457	0.241	0.037	0.352	9.074	9.083
415	-21.448	0.557	-0.530	0.655	0.045	0.571	15.167	14.503
416	------	-----	-----	-----	-----	-----	8.375	8.426
422	22.056	0.241	0.171	0.224	-0.032	0.158	13.709	13.330
504	-6.763	0.224	0.196	0.236	0.115	0.229	1.951	2.462
505	2.699	0.211	-0.098	0.392	0.379	0.088	10.202	10.235
509	16.068	-0.219	0.608	-0.069	0.449	0.019	7.523	7.683
552	9.774	0.022	-0.062	0.356	0.149	0.294	4.531	4.863
553	11.407	0.093	-0.081	0.312	0.342	0.038	5.886	6.100
559	25.437	0.360	0.140	-0.065	0.149	0.082	12.369	12.095
561	-1.729	-0.009	0.414	0.215	0.385	-0.059	2.729	3.212
563	8.552	-0.355	0.043	0.518	-0.039	0.643	6.329	6.561
566	5.184	-0.470	0.299	0.054	0.699	0.277	5.498	5.837
570	-5.184	0.145	0.351	0.166	0.434	-0.089	1.742	2.209
601	3.763	-0.143	0.196	0.698	0.403	-0.227	6.045	6.300
604	8.853	0.115	0.024	0.049	0.403	0.246	7.219	7.416
652	14.373	0.245	-0.030	-0.225	0.147	0.626	7.308	7.438
767	23.317	0.108	0.143	0.208	-0.116	0.341	12.503	12.213
768	------	-----	-----	-----	-----	-----	5.780	6.176
769	-0.323	-0.207	0.360	-0.026	0.586	0.187	1.897	2.325
772	20.473	0.075	-0.025	0.422	0.074	0.104	13.535	13.238
774	9.737	0.168	0.003	0.165	0.197	0.462	17.290	16.785
785	-12.262	0.397	-0.024	0.619	-0.326	0.418	6.255	6.510
786	18.523	0.689	0.057	0.136	0.084	-0.090	14.557	13.934
788	0.127	0.076	0.187	0.435	0.214	0.071	7.598	7.725
795	12.894	0.019	-0.310	0.259	0.466	0.357	9.257	9.250
797	18.766	0.281	0.066	-0.025	0.466	-0.005	12.714	12.471
808	13.522	-0.246	-0.132	0.337	0.651	0.310	12.518	12.250
821	10.812	0.166	0.162	0.126	0.139	0.264	9.674	9.714
822	18.188	0.204	0.014	0.143	0.363	0.052	12.498	12.181

Pooled LSQ	α=8.960	β_1=0.109	β_2=0.144	β_3=0.224	β_4=0.185	β_5=0.189
Pooled LSQ with Adjusted α		β_1=0.103	β_2=0.113	β_3=0.235	β_4=0.230	β_5=0.188
Bayesian Equal Slope Model		β_1=0.103	β_2=0.117	β_3=0.233	β_4=0.226	β_5=0.188

Table 25

School Effectiveness Indices For Sample I Of Cohort I (76-78)

School	Within LSQ (L)	Within LSQ (M)	Within LSQ (H)	Pooled LSQ With Adjusted α	Bayesian Equal Slope Model	Individual Residuals	School Residuals	Dyer's SEIs	Mean Differences
51	40.262	58.576	76.890	58.948	59.174	-1.624	-1.386	3	-2.781
101	42.682	60.520	78.357	60.525	60.586	-0.299	-1.191	3	0.991
102	43.991	60.673	77.355	60.227	60.354	-0.355	-0.182	3	-1.611
106	44.266	60.723	77.179	60.249	60.359	-0.411	-0.542	3	1.022
205	37.230	60.580	83.930	57.998	58.547	-2.570	-1.412	3	-2.762
206	38.796	55.056	71.317	56.909	57.317	-3.929	-5.187	1	-2.201
207	46.220	62.419	78.619	58.528	59.000	-2.020	-2.070	2	-1.812
209	43.977	65.536	87.094	64.445	64.101	3.923	4.013	4	2.121
220	42.627	58.825	75.022	58.581	59.072	-1.771	-1.580	3	-4.569
225	------	------	------	55.945	57.496	-4.739	-5.664	1	-5.635
226	45.238	61.161	77.084	61.362	61.886	1.480	2.515	4	-0.507
235	43.726	59.797	75.867	60.715	60.814	0.325	0.640	3	-1.374
302	43.538	59.223	74.908	55.785	56.665	-4.475	-3.123	2	-5.727
304	34.310	51.545	68.280	52.581	54.290	-8.081	-7.782	1	-9.902
307	------	------	------	56.923	58.318	-3.560	-2.535	2	-2.928
308	47.478	65.547	83.616	66.517	66.041	5.864	6.085	5	5.538
310	39.389	56.915	73.940	57.922	58.270	-2.482	-2.383	2	-4.426
312	47.495	63.340	79.185	66.850	66.118	6.211	6.241	5	6.870
313	44.748	63.149	81.549	63.656	63.501	3.157	3.758	4	1.752
409	------	------	------	53.721	55.786	-6.714	-7.123	1	-8.886
414	39.487	61.145	82.804	63.504	63.181	3.208	3.442	4	1.391
415	50.996	67.019	83.043	67.627	66.381	7.100	7.066	5	5.918
416	------	------	------	65.109	64.110	4.566	3.756	4	3.363
422	58.560	69.225	79.890	65.262	64.675	4.868	5.175	5	2.346
504	35.847	52.756	69.666	54.972	55.694	-5.472	-5.006	1	-8.296
505	39.866	60.558	81.250	62.077	62.076	1.756	2.721	4	-0.587
509	48.513	59.137	69.761	59.034	59.322	-1.370	-0.910	3	-2.767
552	39.917	54.099	68.282	55.380	56.021	-5.335	-5.693	1	-5.778
553	42.579	60.250	77.922	58.726	59.014	-2.079	-2.761	2	-0.923
559	44.306	61.562	78.819	61.494	61.444	0.943	1.237	3	-0.046
561	44.217	57.824	71.431	56.390	57.065	-3.971	-3.947	2	-5.521
563	38.308	57.964	77.119	58.359	58.743	-2.146	-1.383	3	-2.844
566	43.436	57.897	72.359	60.958	60.915	-0.108	-1.561	3	2.993
570	41.191	60.396	79.601	58.993	59.242	-1.554	-2.029	2	-4.091
601	39.655	56.278	72.901	55.312	55.869	-4.742	-2.571	2	-7.949
604	42.946	60.347	77.749	59.840	60.034	-0.667	-0.354	3	-2.308
652	44.186	51.759	79.333	60.822	60.846	0.109	0.286	3	0.987
767	44.681	60.418	76.155	60.985	61.020	0.470	0.906	3	-0.366
768	------	------	------	60.656	60.717	0.033	0.458	3	-0.614
769	40.188	57.172	74.156	56.173	56.701	-4.334	-3.388	2	-5.007
772	53.059	63.526	73.993	65.129	64.672	4.312	3.905	4	5.009
774	48.171	65.341	82.511	66.103	65.544	5.343	4.989	5	5.336
785	55.429	67.220	79.010	64.426	63.913	3.596	2.826	4	3.756
786	53.362	67.421	81.480	66.714	65.714	6.181	6.010	5	5.020
788	42.694	58.215	73.736	59.124	59.315	-1.565	-1.192	3	-0.710
795	37.298	57.357	77.416	58.924	59.217	-1.831	-2.629	2	-1.974
797	42.284	62.641	82.997	64.435	63.998	3.882	3.453	4	2.827
808	50.335	66.031	81.726	65.562	64.929	4.871	4.360	4	4.658
821	47.827	63.050	78.274	62.691	62.581	2.149	2.243	4	0.675
822	44.109	61.782	79.454	60.406	60.507	-0.271	-0.503	3	-0.268

Table 26

School Effectiveness Indices For Sample II Of Cohort I (76-78)

School	Within LSQ (L)	Within LSQ (M)	Within LSQ (H)	Pooled LSQ With Adjusted α	Bayesian Equal Slope Model	Individual Residuals	School Residuals	Dyer's SEIs	Mean Differences
51	41.847	59.038	76.229	60.154	60.079	-0.756	-0.348	3	-1.323
101	41.432	56.040	70.649	56.896	57.202	-4.122	-4.363	2	-3.370
102	34.512	58.209	81.906	59.005	59.016	-1.852	-0.802	3	-2.545
106	45.426	61.665	77.905	62.225	61.969	1.053	0.632	3	3.101
205	48.961	57.457	65.954	59.841	59.825	-1.066	0.323	3	-0.133
206	41.599	58.239	74.879	57.315	57.357	-3.946	-6.033	1	-2.354
207	36.609	59.822	83.035	58.803	58.902	-2.061	-0.871	3	-1.460
209	47.053	62.938	78.824	62.415	62.204	1.490	1.215	3	1.322
220	41.386	54.726	68.067	54.148	54.809	-6.424	-5.015	1	-7.939
225	------	------	------	51.731	53.473	-8.516	-3.005	2	-11.545
226	42.994	60.996	78.998	61.034	60.932	0.243	1.830	4	-0.924
235	42.144	59.996	77.847	60.745	60.653	-0.061	0.194	3	-1.310
302	45.370	58.919	72.468	53.643	54.181	-7.238	-6.560	1	-6.575
304	43.348	57.852	72.355	52.273	53.164	-8.018	-1.834	2	-10.387
307	------	------	------	57.575	58.093	-3.330	-2.085	2	-2.745
308	47.495	64.508	81.520	65.129	64.783	4.117	2.525	4	3.626
310	37.397	56.485	75.573	57.409	57.471	-3.461	-4.770	2	-4.613
312	54.013	67.115	80.218	64.386	65.753	5.188	2.931	4	5.455
313	52.872	67.591	82.310	63.996	66.774	6.244	6.112	5	5.253
409	------	------	------	54.041	55.266	-6.798	-7.161	1	-8.515
414	42.097	60.808	79.519	62.154	61.924	1.183	0.083	3	-0.142
415	37.328	60.295	83.262	68.247	67.344	7.597	9.124	5	6.038
416	------	------	------	61.455	61.267	0.648	0.062	3	-1.453
422	53.760	68.415	83.070	66.789	66.171	5.912	5.208	5	4.056
504	34.952	54.104	73.257	55.031	55.303	-5.619	-4.018	2	-7.877
505	45.489	62.798	80.108	63.286	63.076	2.483	3.582	4	1.471
509	46.476	62.383	78.290	60.603	60.524	-0.144	1.820	4	-0.573
552	43.175	56.799	70.422	57.611	57.704	-3.305	-2.254	2	-3.003
553	42.652	54.979	67.305	58.966	58.941	-2.366	-2.975	2	0.776
559	52.331	65.709	79.086	65.449	64.936	4.481	4.500	4	4.779
561	37.095	55.056	73.018	55.809	56.053	-5.165	-5.844	1	-5.402
563	44.116	58.678	73.239	59.409	59.402	-1.586	-1.895	2	-1.879
566	40.753	56.946	73.139	58.578	58.678	-2.442	-1.510	3	-1.623
570	35.987	55.166	74.344	54.822	55.050	-6.025	-6.300	1	-7.528
601	44.732	60.417	76.101	59.125	59.141	-1.371	1.781	4	-3.518
604	44.185	60.076	75.966	60.299	60.257	-0.321	1.895	4	-2.078
652	45.427	61.127	76.827	60.388	60.279	-0.879	-2.254	2	0.832
767	51.767	65.046	78.325	65.583	65.054	4.516	2.718	4	4.369
768	------	------	------	58.860	59.017	-1.996	1.201	3	-1.232
769	36.161	53.664	71.167	54.977	55.166	-5.991	-5.320	1	-6.056
772	49.342	60.720	72.097	66.615	66.079	5.294	4.218	4	8.086
774	51.920	70.381	89.842	70.370	69.626	9.105	5.520	5	10.116
785	34.847	54.839	74.831	59.335	59.351	-1.637	-3.592	2	-2.295
786	55.147	72.006	88.865	67.637	66.775	6.788	7.915	5	7.479
788	42.127	60.104	78.082	60.678	60.566	-0.369	-1.414	3	-0.496
795	48.715	62.520	76.325	62.337	62.091	1.331	0.982	3	1.626
797	51.414	66.398	81.382	65.794	65.312	5.132	7.347	5	3.798
808	54.526	70.559	86.592	65.598	65.091	4.369	1.745	4	4.698
821	46.340	62.945	79.549	62.754	62.555	1.988	3.548	4	1.115
822	51.334	65.729	80.123	65.578	65.022	4.283	1.208	3	5.671

Two measures of the stability of school effectiveness indices across samples were examined. The first measure was obtained by calculating the correlation between the indices for the two random samples. Table 27 shows the correlations between subsample effectiveness indices. The magnitude of the correlations between effectiveness indices for one subsample and the indices for a second subsample was reasonably high. Stability of subsample indices was highest for indices calculated using pooled least squares regression (r=.85), Bayesian m-group equal slope regression (r=.84), pooled least squares regression with adjusted alpha (r=.84), and mean differences (r=.83). Indices calculated from school residuals and Dyer's SEIs produced slightly less stable results (r=.77 and r=.75, respectively). As expected, the indices based on the within least squares regression model were less stable than the other indices. In addition, those indices for low- and high-scoring students were less stable than those involving mean-scoring students. These results are consistent with those results reported by Dyer, Linn and Patton (1969).

In addition, reliability coefficients for the school effectiveness indices were computed by means of analysis of variance. Using school effectiveness indices from both random halves, the variation of a particular set of school in-

Table 27

Correlations Between School Effectiveness Indices
for Sample I and Sample II of Cohort I (76-78)

School Effectiveness Index	Correlation Coefficient
1. Within LSQ (L)	0.32
2. Within LSQ (M)	0.62
3. Within LSQ (H)	0.49
4. Pooled LSQ with Adjusted α	0.84
5. Bayesian M-Group Equal Slope	0.84
6. Individual Residuals	0.85
7. School Residuals	0.77
8. Dyer's SEIs	0.75
9. Mean Differences	0.83

Note. Correlations based on n=50 except for the first three indices where n=45.

dices was divided into among-school and among-sample variation. The reliability of the mean of school effectiveness indices for the two samples, was then calculated using Winer's (1962) formula. Variance components and reliability coefficients for each school effectiveness indices are found in Table 28. These reliability coefficients, although slightly higher than the correlations of subsample indices, exhibit essentially the same pattern as the first measure of stability across samples. These results are very similar to those found by Marco (1974). However, the stability of indices for low-scoring students (r=.49) was considerably less than that found by Marco (r=.77).

STABILITY OF SCHOOL EFFECTIVENESS INDICES ACROSS CONSECUTIVE CLASSES

Regression analyses were performed independently on two sets of matched longitudinal data, Cohort I (76-78) and Cohort II (77-79). Nine school effectiveness indices were computed for each school using each cohort independently. Correlation coefficients between school effectiveness indices for Cohort I and Cohort II were used as measures of the stability of these indices across consecutive classes. Table 29 gives the values of the correlations between sets of indices for two consecutive classes.

Table 23

Variance Components and Reliability Coefficients for School Effectiveness Indices for Cohort I (76-78)

School Effectiveness Index	Variance Component		Reliability Coefficient
	σ^2schools	σ^2samples	
1. Within LSQ (L)	9.76	20.22	0.49
2. Within LSQ (M)	11.36	7.22	0.76
3. Within LSQ (H)	11.57	11.94	0.66
4. Pooled LSQ with Adjusted α	16.84	2.75	0.92
5. Bayesian M-Group Equal Slope	13.34	2.38	0.92
6. Individual Residuals	14.12	2.67	0.91
7. School Residuals	14.70	3.34	0.90
8. Dyer's SEIs	1.11	0.36	0.86
9. Mean Differences	16.53	3.74	0.90

Note. Analysis based on n=50 except for the first three indices where n=45.

Table 29

Correlations Between School Effectiveness Indices for Cohort I (76-78) and Cohort II (77-79)

School Effectiveness Index	Correlation Coefficient
1. Within LSQ (L)	0.38
2. Within LSQ (M)	0.42
3. Within LSQ (H)	0.28
4. Pooled LSQ with Adjusted α	0.44
5. Bayesian M-Group Equal Slope	0.44
6. Individual Residuals	0.44
7. School Residuals	0.47
8. Dyer's SEIs	0.40
9. Mean Differences	0.37

Note. Correlations based on n=50.

The correlations between indices of consecutive classes were relatively consistant across all nine effectiveness indices. The correlations ranged from .28 for high-scoring student indices to .47 for school residual indices. These results are consistent with those of Gastright (1974) and Acland (1972). Gastright reported consecutive class correlations based on residuals from regression analyses based on school unit data ranging from .25 to .56 while Acland reported correlations among residuals for consecutive classes using unmatched student groups around .40. Forsyth's (1973) correlations between residuals for consecutive years were somewhat lower (median r=.28).

Chapter V

CONCLUSIONS AND IMPLICATIONS

OVERVIEW

This research study empirically investigated four sets of questions arising out of the development of reliable measures of school effectiveness. Initial questions focused on the predictive accuracy of classical and Bayesian regression models using individual student information as predictors. A second set of questions examined the comparability of school effectiveness indices obtained using classical and Bayesian regression models. The third and fourth sets of questions addressed the issue of stability of these indices across samples and across consecutive classes. The research was conducted in context of several important factors which may affect the interpretation of the study's results.

Convey (1974) indicated that in conducting school effectiveness studies several assumptions must be made explicit. First, the researcher must assume that real differences do exist in terms of effectiveness from school to school along at least one dimension. If differences between schools are observed after application of one of the models, "at least part of the difference is due to differential effectiveness and not merely to artifacts of the statistical method employed (Convey, 1974, p. 3)."

A second assumption is that given a dimension along which schools are differentially effective, the output measure used in the statistical analyses adequately represents this dimension. In this study, it is assumed that the composite NCE score of the ITBS adequately represents academic achievement and that schools can be ranked on this basis.

A third assumption of school effectiveness studies deals with the attribution of differences in student outcomes to measurable differences in school variables. If differential effects of schools cannot be attributed to educational process variables but only reflect student characteristics or surrounding conditions for which the school staff has little or no control over, then school effects may not exist.

Finally, it is assumed that all schools within this study have a common objective to increase the basic academic skills of their students. If emphasis of this objective varies among schools, any attempt to compare schools may lead to the obvious conclusion that schools that do not place an emphasis on increasing basic skills will appear to be ineffective.

LIMITATIONS OF THE STUDY

In addition to the numerous assumptions that must be implicitly made when undertaking a school effectiveness study, there exist several methodological limitations in this study. The first limitation deals with the selection of predictor variables for each of the models included in the study. Noticeably missing from inclusion in each model are variables which reflect latent traits of the students (i.e., IQ and motivational factors) as well as environmental variables (i.e., home and community factors). Guthrie (1970) in his review of 19 school effectiveness studies found SES to be strongly related to achievement. If these types of variables entered into the analysis, different variables may have emerged as important. However, the inclusion of these measures into the models would have made it necessary to administer additional tests and quantify environmental factors. This was not practical. For this study, it was assumed that the relevant preexisting conditions were adequately reflected in the differences of pretest scores.

A second limitation in this study was the use of Shigemasu's equal slope model. If the Bayesian m-group regression method was used without the equal slope assumption different results may have resulted. If the true relationship between the pretest subtest scores and posttest composite scores differ substantially across the m groups, then using the equal slope assumption may provide tenous results.

SUMMARY OF CONCLUSIONS

The following are the general conclusions related to the major questions investigated in this study.

To what extent are classical and Bayesian regression models comparable when applied to individual, matched longitudinal samples? How comparable is the predictive efficiency of these regression models?

The estimated regression coefficients obtained by the Bayesian m-group equal slope model and the pooled least squares regression with adjusted alphas model were almost identical. Three of the five estimated regression coefficients obtained using the pooled least squares regression model compared favorably with the Bayesian m-group equal slope model. The coefficients for the within-school least squares regression model varied greatly with many coefficients having negative weights. These negative regression coefficient estimates probably are due to sampling variation and are merely reflecting idiosyncrasies found in the data.

Examination of mean squared errors, mean absolute errors, and correlations between observed and predicted scores as criteria for comparing predictive efficiency revealed that the within-school least squares regression model minimized the error for each school. The pooled least squares regression model allowed for the most error while the Bayesian m-group equal slope model and the pooled least squares regression with adjusted alpha model produced mean squared errors, mean absolute errors, and correlations between these two methods. This was not surprising since the Bayesian m-group equal slope regression represents a weighted average of a particular school's classical regression line and the regression line obtained for all schools collectively.

Cross-validation analyses revealed the error of prediction, on the average, was considerably greater for the withinschool least squares regression method when compared to the remaining three methods. The cross-validation analysis confirmed the results of Lissitz and Schoenfeldt (1974) with some improvement in prediction obtained by the Bayesian m-group analyis over the within-school least squares regression, but no significant improvement was offered when compared to the pooled least squares regression method.

In summary, it appears that the within-school least squares regression method is substantially different from the remaining three models. In terms of comparability of estimated regression coefficients, the within-school least squares regression model tended to reflect sampling fluctuations and idiosyncrasies found within each school's data. In terms of predictive efficiency, the within-school least squares regression method is somewhat inferior to the remaining three models. The Bayesian m-group equal slope model, adjusted and nonadjusted pooled least squares models tend to moderate extreme values. It appears that the predictive accuracy of the Bayesian m-group equal slope model shows no appreciable advantage over the pooled least squares regression with adjusted alphas model or the pooled least squares regression model with no adjustments.

To what extent are school effectiveness indices comparable when obtained by each of six different methods?

The indices produced by all six methods appear to be capable of representing the relative effectiveness of the schools involved in the study. All six methods yielded indices that are highly correlated. Correlations were generally above .85 with the exception of those involving indices produced for low- and high-scoring students using within-school regression methods. Although the effectiveness indices do differ from each other, it is not known which of these school effectiveness indices best estimate "true" school effectiveness.

This study has shown that school effectiveness indices for low-, middle-, and high-scoring students differed substantially. The correlation of the school effectiveness indices for low-scoring students with the indices for high-scoring students was only .51. This relationship is consistent with the findings of Marco (1974) and raises some doubts about using one index to measure the effectiveness of a school for a given group of students.

Indirect comparisons of school effectiveness indices based on their correlational patterns with additional school variables provided very little additional information for interpreting these indices. Only two variables, average pupil attendance and property destruction, were correlated with the school effectiveness indices. Correlational patterns between school effectiveness indices for these school variables did not differ substantially among indices.

<u>How stable are these measures of school effectiveness across samples of students</u>?

Two measures of the stability of school effectivenss indices across random samples were obtained in this study. Correlations between the school effectiveness indices obtained for two independent samples served as the first measure of stability. Reliability coefficients were also obtained for the school effectiveness indices by means of analysis of variance techniques. In both cases all but those indices obtained by the within-school least squares regression model produced indices that appeared to be stable enough warrant use as measures of school effectiveness.

The correlation between effectiveness indices of the two subsamples ranged from .32 to .62 for those indices based on within-school least squares regression methods, and between .75 and .85 for the remaining indices. Coefficients obtained using analysis of variance techniques produced slightly higher measures but exhibited essentially the same pattern. It appears that school effectiveness indices remain relatively stable from one subsample to another subsample within a given year when the methods used to calculate these indices moderate the importance of extreme values. These results are consistent with the findings reported by Dyer, Linn and Patton (1969) and Marco (1974).

<u>How stable are these measures of school effectiveness for consecutive classes?</u>

School effectiveness indices were relatively unstable from one year to the next. Correlations between school effectiveness indices of consecutive classes were relatively consistent across all effectiveness indices ranging from .28 for the within-school least squares index for high-scoring students to .47 for indices based on school residuals. These results are consistent with those of Acland (1972) and Gastright (1974) and somewhat higher than those reported by

Forsyth (1973). Stability from class to class were of a much lower magnitude than those values reflecting stability from sample to sample.

It is reasonable to believe that these estimates of stability from class to class represent a lower bound of the true stability indices. Systematic or out-of-the-ordinary changes taking place in the schools participating in this study may have occurred which may explain some of the differences in effectiveness. No check was made for these possible changes in this study. If identification of over- and under-achieving schools is desired, computation of indices for two consecutive classes would seem reasonable. Those schools that are consistently low or consistently high should be identified as outliers and studied for possible causes.

RECOMMENDATIONS FOR CONDUCTING SCHOOL EFFECTIVENESS STUDIES

A primary reason for conducting school effectiveness studies is to be able to identify schools that are relatively effective or ineffective and examine information about school variables that contribute to relative effectiveness. While the findings of this study indicate that each of the models appear capable of producing indices that accurately reflect effectiveness rankings of schools involved, there are a number of considerations that should be examined when undertaking school effectiveness research. The following sections contain recommendations for conducting school effectiveness studies.

<u>The criterion variable</u>

The establishment of a criterion variable which adequately reflects the dimension on which those schools are to be compared in terms of effectiveness is of utmost importance. The specification of the appropriate outcome variable must represent a dimension or objective common to all schools involved in the study. Identification of outcome variables that adequately represent this dimension should be based on theory or previous research. However, all to often, the decision of which criterion variable to use is based on convenience. Since most school systems utilize some kind of standardized testing of basic skills on a regular basis, scores for these tests are commonly used as outcome measures. The decision to use composite scores or a specific subtest score as a criterion variable hinges on the dimension selected to be studied. It seems unlikely that school effectiveness can easily be classified into only one dimension. Effectiveness studies may need to shift to multiple criterion variables in a search for overall effectiveness ratings.

The predictor variables

Once criterion measures have been identified, what input variables should be used? The inclusion of predictors into the model should again be based on theory or previous research. The number of predictors is not as important as the type of predictors. Some measure of the student's entering capabilities should be taken into account. If equivalent forms of standardized tests are used for both input and output measures, and most of the reliable variation of the subtest scores is represented in a composite score, it is possible to replace the subtest scores with one composite variable as a predictor. Reducing the number of achievement variables used as predictors may allow for the inclusion of more nonachievement variables into the prediction equation. If individual student scores are to be used as input variables, the inclusion of some variables measuring latent traits of the pupils or environmental variables may increase the predictability of the model. However, inclusion of these types of variables as predictors may necessitate the administration of additional tests or quantification of factors which may be considered impractical. If school means of standardized tests are used as input variables the inclusion of other school related variables may improve prediction.

Daniel and Grobe (1981) have listed ten factors identified in different research as being important influences of effective schooling. These are (1) principal characteristics, (2) time-related factors, (3) program coordination, (4) teacher characteristics, (5) instructional materials and methods, (6) teacher/student interaction, (7) instructional emphasis, (8) instructional accountability, (9) student background characteristics, and (10) organizational variables. Variables selected as potential predictors should be entered into the model and after examining their predictability some decision about their appropriateness for inclusion should be made.

<u>The model selection</u>

The appropriateness of model specification should be examined prior to utilization in effectiveness studies. "Model misspecification has been shown to produce significant changes in the number of school outliers, school classifications and magnitude of school effects (Conner, 1979, p. 79)." Six models were examined in this study as possible ways of producing school effectiveness indices. Each of the models appear capable of representing the relative effectiveness of the schools involved in the study. It is not known which indices best estimates the "true" school effectiveness since no attempt has been made to validate these indices using schools of known quality. To say that one method is better than another at this time is inappropriate. Each of the models have their advantages and disadvantages as noted in previous research (Harris, 1967; Linn and Slinde, 1977; Convey, 1973).

Examining the correlations between indices produced by these six methods, as well as examining the measures of stability of these indices across samples and across consecutive classes has provided no conclusive evidence to indicate that any one model is superior to the others. By no means has the Bayesian m-group equal slope model set itself apart from the others. In fact, all the models which assumed parallelism of the regression planes based on either individual scores or mean scores performed equally well. To this end, since school means may be more readily available than individual scores, the school residual model may be considered better in a cost effectiveness sense. In addition, Dyer's SEIs are readily calculated from data obtained from this model and should provide adequate discrimination between schools considered effective or ineffective.

The within-school least squares regression model does not require the assumption of homogenity of regression. This model does require computation of a regression surface for each school. Although a single overall effectiveness index for each school cannot be calculated, it does allow for comparison of schools at selected points of interest. Although this model yielded indices which tended to be the least stable of all indices examined in this study, there may be some merit in this model. If schools are suspected of being differentially effective for students possessing different characteristics than the use of this model may be appropriate.

Although mean score differences compared favorably with other school effectiveness indices in this study, their use as the sole indicator of effectiveness is not recommended. If it is used it should be used in conjunction with other indices. One primary criticism of this model is the fundamental assumption that in the absence of the treatment (i.e., schools) absolute gain for all students would be the same (Campbell & Erlebacher, 1970). This assumption is seldom met due to the numerous other variables that have been shown to affect output.

Classification of schools

Once a model has been selected and the analysis completed, a method of classifying schools as effective or ineffective is needed. Convey (1977) examined three models to determine the relative effectiveness of schools: (1) Dyer's SEIs, (2) Sheffe's hyperbolic confidence bands, and (3) Gafarian's linear confidence bands. The results indicated that the extreme SEIs were appropriate for identification of school performing above or below expectation. Discrimination between the three middle categories did not appear to be warranted. Burke (1978) indicated that confidence intervals based on 1.5 and 2.0 standard errors should be considered appropriate criteria for classification. If incorrect identification of outliers is to be avoided a more conservative strategy would be appropriate. Those schools identified using the most conservative estimates should be classified as effective or ineffective. Less conservative intervals may identify additional schools considered tentative outliers. Interpretation of these classifications need to take into account the possibility of incorrect classifications.

Investigation of outliers

School effectiveness studies should not stop at the identification of effective or ineffective schools. School variables which may contribute to the relative effectiveness need to be examined. Prior to comparing those schools identified as effective with those identifed as ineffective, some effort should be made to determine if some non-school variables, that is, those variable for which the school staff has little or no control over, may have caused a school to be classified as an outlier. If no evidence appears that some outside condition was responsible for the classification then examination of differences in school process variables for effective and ineffective schools should take place. It is imperative that accountability decisions not be made solely on the basis of effectiveness indices. The indices provide a measure that, in conjunction with all the other information about the schools, may prove valuable in the identificaiton of effective and ineffective schools.

RECOMMENDATIONS FOR FUTURE RESEARCH

This study has addressed questions relative to stability of indices produced by a number of methods. As a result of these findings, some questions about the models used in the development of school effectiveness indices have been been answered. However some additional questions have also been raised.

Of utmost importance in the research of school effectiveness is the validation of the regression approaches examined in this study to determine if the rankings of the indices obtained from these models are consistent with an external criteria of effectiveness. External criterion may consist of ranks established by a team of expert evaluators or ranks based on school reputations within in the community. Agreement of results would provide some evidence of the validity of the regression models. If the results were inconsistent, the question of which method was deemed more valid would not be answered. To this end, other external validation criteria need to be developed and examined for consistencies.

Secondly, some studies should be conducted which take into account subgroups within a school that may differ substantially in terms of academic performance. To assume that all students learn and respond to the same kinds of educational environments in the same manner may not be appropriate. Not only should groupings be developed based on pretest information (i.e., low-, middle-, and high-scoring students) but also examination of other background characteristics may prove insightful relative to the behavior of effectiveness models under study. Whether the within-school regression model will be a superior alternative in this type of situation remains unanswered.

Thirdly, although the Bayesian m-group regression equal slopes model did not produced superior results in terms of predictive efficiency or stability of its indices between samples and across time, one should not exclude a Bayesian methodology from future school effectiveness research. Use of a Bayesian method without the equal slope restriction may have produced entirely different results. Since the Bayesian model includes within-group least squares, pooled least squares, pooled least squares with adjusted alphas, and a number of other models as special cases, this model seems suitable for continued scrutiny as a plausible model for school effectiveness studies. Its biggest advantage is that it does not commit the user to any one, possibly false model.

Finally, the use of multivariate techniques should be incorporated into more effectiveness studies. Effectiveness of schools needs to be examined on more than one dimension. A single index based on a single criterion does not seem appropriate. Multiple dependent measures each providing information for a unique dimension on which a school can be identified as effective or ineffective needs to be examined. Applying the models used in this study to different criterion variables each time would provide an effectiveness profile. Averaging the indices produced for each dependent variable may provide an overall effectiveness index. However, the use of multivariate multiple linear regression techniques seems preferable. All dependent variables would be considered simultaneously while taking into consideration the inherent dependency present among the measures.

BIBLIOGRAPHY

Acland, H. The impact of schooling in the elementary grades, The Center of Educational Policy Research, Harvard Graduate School of Education, 1972.

Bolt, R. F. Comparison of a Bayesian and a least squares method of education predictions, Research Bulletin 75-15, Princeton, N. J.: Educational Testing Service, 1975.

Burk, K. A comparison of four regression models for predicting school mean test scores in school effectiveness studies. Unpublished doctoral dissertation, Florida State University, 1978.

California Test Bureau/McGraw-Hill. Technical Report: Comprehensive Tests of Basic Skills. Monterey: McGraw-Hill, 1970.

Campbell, D. T. & Erlebacher, A. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Helmuth (Ed.), Disadvantaged child: Compensatory education a national debate (Vol. 3). New York: Brunner/Magel, 1970.

Coffman, W. E. & Shigemasu, K. Appraising school effectiveness using a Bayesian method. A paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978. (ERIC Document Reproduction Service No. ED 159 220)

Connor, E. G. The effect of regression model specification on school effectiveness indices and size of school effects. Unpublished doctoral dissertation, Florida State University, 1979.

Convey, J. J. Some methodological considerations for school effectiveness studies. Paper presented at the annual meeting of the Florida Educational Research Association, Tallahassee, Florida, January, 1974. (ERIC Document Reproduction Service No. ED 092 597)

Convey, J. J. A validation of three models for producing school indices. Paper presented at the annual meeting of American Educational Research Association, Washington, March 1975. (ERIC Document Reproduction Service No. ED 106 368)

Convey, J. J. Determining school effectiveness following a regression analysis, Journal of Educational Statistics, 1977, 2, (1), 29-39.

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

Daniel, G. S. & Grobe, R. P. Variables associated with effective schooling. September, 1981. (ERIC Document Reproduction Service No. ED 207 221)

DeFinetti, B., Foresight: Its logical laws, its subjective sources. In H. E. Kyburg & H. E. Smokler (Eds.), Studies in Subjective Probability. New York: Wiley, 1964.

Dyer, H. S. The Pennsylvania plan. Science Education. 1966, 50, 242-248.

Dyer, H. S. Toward objective criteria of professional accountability in the schools of New York City. Phi Delta Kappan, 1970, 52, 206-211.

Dyer, H. S., Linn, R. L., & Patton, M. J. Feasibility study of educational performance indicators: Final report to New York State Education Department. Princeton, N. J.: Educational Testing Service, 1967. (ERIC Document Reproduction Service No. ED 079 353)

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 based on achievement tests. American Educational Research Journal, 1969, 4, 591-605.

Fillos, R. M. & Bowman, C. M. Examining school effects on achievement: A conceptual model and an example. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1976. (ERIC Document Reproduction Service No. ED 122 387)

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.

Frederikson, J. R. Models for determining school effectiveness. A paper presented at the annual meeting of American Educational Research Association, Boston, April, 1980. (ERIC Document Reproduction Service No. ED 189 149)

Gastright, J. F. Some empirical evidence on the comparability of school unit residuals based on achievement and non-achievement variables. Paper presented at the annual meeting of the American Educational Research Association, New York, April 1977. (ERIC Document Reproduction Service No. ED 135 850)

Gastright, J. F. Some empirical evidence on the stability of discrepancy measures based on observed and predicted school means on achievement tests. Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1974. (ERIC Document Reproduction Service No. ED 093 957)

Guthrie, J. W. A survey of school effectiveness studies. In U. S. Department of Health, Education, and Welfare, Do teachers make a difference? Washington, D. C.: U. S. Government Printing Office, 1970.

Hanushek, E. A. The production of education, teacher quality, and efficiency. In U. S. Department of Health, Education, and Welfare, Do teachers make a difference? Washington, D. C.: U. S. Government Printing Office, 1970.

Henriksen, L. W. A Bayesian prediction of an adjusted criteria in the prediction of success of particular areas of study at the Florida State University. Unpublished doctoral dissertation, Florida State University, 1973.

Hilton, T. L., & Patrick, C. Cross-sectional versus longitudinal data: An empirical comparison of mean differences on academic growth. Journal of Educational Measurement, 1970, 1, 15-24.

Hinkle, D. E., & Polhamus, E. C. Bayesian m-group regression applied in a community college developmental studies program. Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980. (ERIC Document Reproduction Service No. ED 187 747)

Houston, C. A., Jr. A comparative study of the effectiveness of two Bayesian models for predicting the academic success of selected allied health students enrolled in the comprehensive community college. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, 1976.

Jackson, P. M., Novick, M. R., & Thayer, D. T. Estimating regressions on m-groups. The Bristish Journal of Mathematical Psychology, 1971 24, 129-153.

Jones, P. K., & Novick, M. R. Implementation of a Bayesian system for prediction in m-groups. ACT Technical Bulletin No. 13. Iowa City, Iowa: The American College Testing Program, 1972. (ERIC Document Reproduction Service No. ED 080 589)

Kelley, T. L. Fundamentals of Statistics. Cambridge: Harvard University Press, 1927.

Laughlin, J. E. Bayesian inference in regression. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign, 1976.

Lindley, D. V. A Bayesian solution for some educational prediction problems: II. Research Bulletin 69-91. Princeton, N.J.: Educational Testing Service, 1969.

Lindley, D. V. A Bayesian solution for some educational prediction problems: III. Research Bulletin 70-133. Princeton, N.J.: Educational Testing Service, 1970.

Linn, R., Werts, C. & Tucker, L. The interpretation of regression coefficients in a school effects model. Educational and Psychological Measurement, 1971, 31, 85-93.

Lissitz, R. W. & Schoenfeldt, L. F. Moderator subgroups for the estimation of educational performance: A comparison of prediction models. American Educational Research Journal, 1974, 11, 63-75.

Marco, G. L. A comparison of selected school effectiveness measures based on longitudinal data. _Journal of Educational Measurement_, 1974, _11_, 225-234.

Marco, G. L., Murphy, R., & Quirk, T. A classification of methods of using student data to assess school effectiveness. _Journal of Educational Measurement_, 1976, _13_, 243-252.

Maryland State Department of Education. _Maryland accountability program report, school year 1973-1974_. Rockville, Md.: Westeast Research, 1975. (ERIC Document Reproduction Service No. ED 118 637)

Maryland State Department of Education. _Maryland accountability program report--year 2, school year 1974-1975_. Rockville, Md.: Westeast Research, 1976. (ERIC Document Reproduction Service No. ED 118 638)

New York State Education Department. _Performance indicators workbook: Edition 1, for calculating school district performance in elementary school reading and arithmetic_. Albany: New York State Education Department, 1973. (ERIC Document Reproduction Service No. ED 080 591)

Novick, M. R., Jackson, P. H. & Thayer, D. T. Bayesian inference and the classical test theory model: Reliability and true scores. _Psychometrica_, 1971, _36_, 261-288.

Novick, M .R. Jackson, P. H., Thayer, D. T., & Cole, N. S. Applications of Bayesian methods to the prediction of educational performance. _ACT Research Report No. 42_., Iowa City, Iowa: The American College Testing Program, 1971.

Novick, M. R., Jackson, P. H., Thayer, D. T. & Cole, N. S. Estimating multiple regression in m groups: A cross-validation study. _British Journal of Mathematical and Statistical Psychology_, 1972, _25_, 33-50.

Novick, M. R. & Jackson, P. H. Further cross-validation analysis of the Bayesian m-group regression method. _American Educational Research Journal_, 1974, _11_, 77-85.

Novick, M. R. Moderator subgroups and Bayesian m-group regression: Some concluding remarks. _American Educational Research Journal_, 1974, _11_, 91-92.

Novick, M. R. & Jackson, P. H. Statistical Methods for Educational and Psychological Research. New York: McGraw-Hill, 1974.

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

Schoenfeldt, L. F., & Lissitz, R. W. Moderator subgroups and Bayesian m-group regression: Some further comments. American Educational Research Journal, 1974, 11, 87-90.

Shigemasu, K. Development validation of a simplified m-group regression model. Journal of Educational Statistics, 1976, (2), 157-180.

Shigemasu, K. Two educational applications of a general linear model. Unpublished doctoral dissertation, University of Iowa, 1974.

Soar, R. S. Accountability: Problems and possibilities in accountability as the measurement of pupils. Paper presented at the annual meeting of the American Educational Research Association, Chicago, 1973.

Werts, C. E., & Linn, R. L. Analyzing school effects: ANCOVA with a fallible covariate. Educational and Psychological Measurement, 1971, 31, 95-104.

Winer, B. J. Statistical principles in experimental design. (2nd ed.) New York: McGraw-Hill, 1971.

The vita has been removed from the scanned document