1971 AERA SIG Business Meeting

Sam Houston presided and indicated he had attended a meeting of all SIG presidents. Annual reports must be submitted to AERA and Bill Connett, our new secretary, has already mailed ours in. Bill reports 32 persons have paid their 1971-72 dues and our treasury shows a balance of $95.00. Keith McNeil was elected president for the coming year and Joe Ward accepted the appointment of symposium organizer. Much discussion centered around Viewpoints and around the convention symposium. At this point we are considering running the symposium in 1972 as we did in 1971. Any suggestions you might have, send to Joe Ward.

Several persons indicated that they were looking for Viewpoints to present ideas rather than full fledged papers. The general quality of the Viewpoints material was questioned. The reactions that I have to these views seems to be similar to most of the groups---Viewpoints is our dissemination vehicle, let's make it what we want. I would encourage members to flood John Williams, Viewpoints editor, with ideas, questions, teaching suggestions critiques of published material, etc. The ideal length of the material would be a single page. That way, we can get more flow of information. Remember, send $1 with every typed page to John Williams.

Keith McNeil
SIG-MLR President 1971-72

P.S. I have a couple of notions in this issue of Viewpoints. I would sure appreciate some reactions, both positive and negative.

ON THE UNIT VECTOR

Often times have trouble explaining the inclusion of the unit vector? Here are some notions that I've tried. With categorical predictor vectors the unit vector is redundant, but I define it as a "1" if the S is in the sample, "0" otherwise. Of course, all the Ss are in the sample, hence all ones in the vector.

With continuous predictor variables the explanation becomes a little more difficult for the student and instructor. The weight for the unit vector represents the "Y-intercept" for the line of best fit, or the plane of best fit, etc. Reference to the weighting coefficient as a "constant" which must be added to each S's score is also quite appropriate.

We've been on a curvilinear binge lately and I came across another possible way of thinking of the unit vector when presenting the notions of curve fitting to my class. In discussing a fourth degree curve, I had the following model on the board:

$$Y_1 = a_oU + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + E_1$$

The model can be better understood in non-standard symbology:

$$Y_1 = a_oU + a_1X_1 + a_2X_1^2 + a_3X_1^3 + a_4X_1^4 + E_1$$

I was explaining that the multiple linear regression procedure could reflect as curvy a line as the investigator so desired. I emphasized the pattern, explaining that if they wanted the fifth degree all they had to do was to add $a_5X_1^5$ to the above model.

For some reason I started from right to left indicating, 4, 3, 2, and then I raised the second term (X_1) to the first power.

Being systematic, I then thought that the first term could be represented as X_1^o rather than U, since any number raised to the 0 power is 1, thus the model can also be represented as:

$$Y_1 = a_o X_1^{o} + a_1 X_1^{1} + a_2 X_1^{2} + a_3 X_1^{3} + a_4 X_1^{4} + E_1$$

Notice that the subscripts of the weighting coefficients correspond to the power. I always wondered why the system said to use a_o as the weighting coefficient for the unit vector. Any comments?

On Attenuating a Multiple R

At one of the sessions at the 1971 AERA, all of the presenters were careful to shrink their R values, and then correct them for attenuation, thus acknowledging the unreliability of the criterion. I entirely agree with the desire to indicate the degree of overfitting of the data, although I would argue for an empirical demonstration rather than application of an estimation formula.

I am not familiar with attenuating R values, and upon a little reflection, I don't believe that I can support that procedure. It seems a little ironic to boast about the degree of predictability one would have, if one had a perfectly reliable criterion. It is probably the case that we will never have perfectly reliable criteria, and furthermore, when one tries to predict a criterion, one must have some faith in the measurement of that criterion. Prediction of a criterion is an exercise in validity, and increasing the "statistical" reliability of the criterion may decrease the validity of the functional relationship.

Statistical analyses should deal with real-world problems. Overfitting of the data is a real-world problem and hence some solution is called for (preferably a real world solution such as cross validation). Perfectly reliable criteria only occur in a dream-world, and hence no solution is called for. Any comments?

April 5, 1971

Dear Significant Others,

After spending an enjoyable and informative weekend with our illustrious, or is that infamous, president, Keith McNeil; a question has been raised to which even we could not find an answer, in the span of one weekend that is (undoubtedly we could have resolved the problem with a few more minutes time). The question is actually mine, being that it is immediately relevant to the research I am conducting. I am involved with Tom Jordan, a fellow Significant other, in a longitudinal preschool development program. Many of our variables contain missing data. In terms of MLR analysis, which is our bread and butter analysis procedure, we have been forced to deal with only those subjects with complete data. Now to the question. Can a correlation matrix, generated by a missing data descriptive program, be utilized in the regression analysis program? If not, why not?

Let me go on and explain my reasoning as to why I should be able to use a correlation matrix based on missing data. My intention in using MLR is not only to explain my existing data but to generalize beyond the range of my data and sample. When I use only those subjects who possess complete information, I am limiting the range on the variables under analysis and limiting my sample. If I use all subjects available and generate my correlation matrix on all the available information, I feel I would be generating a closer approximation of the intercorrelations of a sample the same size but with complete data. It seems to me that in "natural setting" research, where missing data is the price we pay to observe reality, that an analysis of interrelationships based on missing data is reality. How say you?

An afterthought of mine is to include a vector in my models of the number of pieces of datum missing for a subject. Would this solve the dilemma? If it does, which intercorrelation matrix do I use; the missing data matrix or the complete data matrix.

Sincerely,

Steve Spaner

Steve Spaner
Research Associate
CEMREL-EDAP
10646 St. Charles Rock Rd.
St. Ann, Mo. 63074

SS:jb

A REGRESSION/PRINCIPAL COMPONENTS ANALYSIS OF SCHOOL OUTPUTS

by

William L. Duff, Jr.
Samuel R. Houston
Sheldon Bloom[1]

The objective of this study is to identify the correlates of student performance and teacher retention in an inner-city elementary school district. The purpose is to provide urban school administrators with information necessary to cope with the special problems they face in organizing and administering their educational resources.

The study is divided into two parts: a descriptive section and an analytic section. In the descriptive section the writers are concerned with describing the inner urban school system. Here the data to be analyzed are presented and classical regression techniques are used to specify the three basic teacher retention and student performance models. In the second section the data are further analyzed in terms of the unique contribution of a priori specified subsets of predictor variables. This section ends with a comparison of a principal component regression approach to the *a priori* grouping of predictors used in the unique analysis.

SECTION I

Data Description

Vectors selected from a 32 X 128 data matrix, descriptive of the students, the faculty, and the school, are used to specify each of the student performance and teacher retention models. The data are descriptive of elementary schools in Washington, D.C. Public School systems. The data were gathered from census tracts, school records and site visits to the various elementary schools included in the study.[2]

---

[1]Dr. William L. Duff, Jr., Director, Bureau of Business and Public Research, University of Northern Colorado; Dr. Samuel R. Houston, Associate Professor of Statistics and Research Methodology, University of Northern Colorado; Dr. Sheldon Bloom, U.S. Department of Labor, Washington, D.C.

[2]The data were originally gathered by Professor George Carey, Geography Department, Columbia University, for use in "The Passow Report," for the Washington, D.C. Public Schools. After preparation of the report Dr. Carey permitted the authors to use the data.

TABLE 1

Variable Description[3]

Var.No.	Description
1	Percent white (s)
2	Pupil/teacher ratio (pf)
3	Percent married (t)
4	Percent with school-age children (t)
5	Percent under 40 years of age (t)
6	Percent raised in D.C. (t)
7	Percent raised outside D.C., but in the South (t)
8	Percent raised in the South (including D.C.) (t)
9	Percent raised in town of more than 10,000 people (t)
10	Percent raised on a farm (t)
11	Percent reporting parents' income in upper one-half of community (t)
12	Percent male (t)
13	Percent Negro (t)
14	Percent permanent teachers (t)
15	Percent probationary teachers (t)
16	Percent temporary teachers (t)
17	Percent with bachelor's degree (highest degree) (t)
18	Percent with master's degree (t)
19	Number with school-age children in D.C. public school, compared to the number with school-age children (t)
20	Median family income (s)
21	Median years of education of parents (s)
22	Attendance as a percent of enrollment (pf)
23	Ratio, capacity to enrollment (the larger the value, the more space available) (pf)
24	Years experience at present school
25	Years experience in D.C. public school system (t)
26	Total years teaching experience (t)
27	Age of school building (pf)
28	Date of latest addition (pf)
29	Number of classrooms (pf)
30	Number of amenities (pf)
31	Number of substandard facilities (pf)
32	6th grade reading scores (s)
33*	Experience prior to D.C. (t)

* generated variable (var 33 = var 26 - var 25)

[3] In the variable description, (s) = student, (pf) = school physical facilities, (t) = teacher.

TABLE #2

Intercorrelation Matrix for Complete Data Set

	1	2	3	4	5	6	7	8	9	10
1	1.00									
2	-0.23	1.00								
3	0.01	0.11	1.00							
4	-0.31	0.16	0.42	1.00						
5	-0.43	0.18	0.19	0.16	1.00					
6	0.16	-0.07	0.35	0.27	-0.13	1.00				
7	-0.38	0.04	0.13	0.21	0.52	-0.36	1.00			
8	-0.18	-0.02	0.43	0.43	0.33	0.59	0.54	1.00		
9	0.10	-0.17	0.30	0.19	0.12	0.51	-0.19	0.30	1.00	
10	0.02	0.04	0.11	0.28	-0.01	-0.10	0.26	0.13	-0.25	1.00
11	0.44	-0.04	0.26	0.10	0.00	0.12	-0.07	0.05	0.29	0.02
12	-0.14	-0.16	-0.00	0.07	0.14	-0.10	0.14	0.03	-0.05	-0.12
13	-0.89	0.09	-0.04	0.31	0.42	-0.11	0.44	0.28	-0.08	-0.02
14	0.18	-0.23	0.11	0.20	-0.43	0.56	-0.37	0.19	-0.09	-0.12
15	-0.06	0.05	0.24	0.03	0.36	-0.17	0.39	0.19	-0.09	0.32
16	-0.18	0.23	-0.14	-0.21	0.41	-0.56	0.34	-0.22	-0.37	0.09
17	-0.18	0.00	0.15	0.03	0.46	-0.29	0.42	0.11	-0.08	0.24
18	0.04	-0.18	-0.01	-0.04	0.01	0.22	-0.10	0.11	0.11	-0.08
19	-0.24	0.13	0.10	0.31	0.21	0.12	0.11	0.21	-0.01	-0.01
20	0.81	-0.12	-0.06	-0.26	-0.48	0.16	-0.44	-0.24	0.15	-0.02
21	0.63	-0.03	0.04	-0.10	-0.33	0.21	-0.34	-0.11	0.21	0.00
22	0.13	0.23	0.21	0.13	-0.03	0.12	-0.09	0.03	0.28	-0.07
23	0.04	-0.73	0.02	-0.10	-0.17	0.11	0.00	0.10	0.08	0.16
24	-0.10	-0.13	0.08	0.23	-0.22	0.25	-0.00	0.23	0.10	0.00
25	0.32	-0.24	0.11	0.15	-0.58	0.43	-0.37	0.07	0.18	0.01
26	0.34	-0.25	0.04	0.05	-0.71	0.30	-0.37	-0.05	0.05	0.03
27	0.04	-0.19	-0.10	-0.10	0.06	0.00	-0.02	-0.01	-0.03	0.11
28	-0.16	0.32	0.15	0.22	-0.04	-0.02	0.04	0.02	0.10	-0.12
29	0.36	0.43	0.09	0.24	0.13	-0.07	0.20	0.11	-0.01	-0.06
30	-0.23	0.10	0.12	0.26	0.10	0.04	0.13	0.15	0.09	-0.01
31	-0.22	0.26	0.03	0.02	-0.01	-0.06	0.00	-0.06	-0.03	-0.09
32	0.81	-0.10	0.10	-0.20	-0.44	0.22	-0.40	-0.14	0.16	0.03

	11	12	13	14	15	16	17	18	19	20
11	1.00									
12	-0.09	1.00								
13	-0.43	0.21	1.00							
14	0.05	-0.15	-0.09	1.00						
15	0.04	0.01	0.09	-0.40	1.00					
16	-0.06	0.16	0.08	-0.99	0.30	1.00				
17	0.03	0.10	0.23	-0.49	0.72	0.43	1.00			
18	0.01	-0.03	-0.04	0.28	-0.01	-0.29	-0.30	1.00		
19	0.08	-0.04	0.20	0.11	0.06	-0.12	0.01	0.14	1.00	
20	0.41	-0.21	-0.79	0.26	-0.13	-0.25	-0.27	0.09	-0.18	1.00
21	0.43	-0.21	-0.63	0.23	-0.16	-0.22	-0.32	0.12	00.12	0.78
22	0.22	-0.16	-0.16	0.10	-0.13	-0.09	-0.14	0.10	-0.08	0.36
23	0.24	0.01	-0.30	0.18	0.08	-0.19	0.09	0.10	-0.18	0.29
24	-0.18	-0.02	0.21	0.36	-0.16	-0.36	-0.09	0.03	0.11	-0.13
25	0.11	-0.09	-0.24	0.67	-0.41	-0.65	-0.46	0.17	0.02	0.34
26	0.06	-0.08	-0.29	0.62	-0.43	-0.59	-0.46	0.14	-0.07	0.36
27	0.01	-0.07	0.00	0.04	0.13	-0.06	0.23	-0.06	-0.03	0.01
28	-0.02	0.02	0.09	-0.04	-0.07	0.05	-0.17	0.04	0.11	-0.06
29	-0.10	0.04	0.29	-0.12	-0.04	0.13	-0.11	0.05	0.19	-0.29
30	-0.16	0.08	0.19	-0.03	0.04	0.03	-0.06	0.25	0.15	-0.20
31	-0.02	-0.09	0.02	-0.06	-0.07	0.07	-0.17	0.12	0.20	-0.08
32	0.35	-0.21	-0.76	0.25	-0.09	-0.25	-0.26	0.11	-0.19	0.80

TABLE # 2 (cont.)

	21	22	23	24	25	26	27	28	29	30
21	1.00									
22	0.46	1.00								
23	0.30	0.03	1.00							
24	-0.16	0.07	-0.06	1.00						
25	0.28	0.18	0.23	0.47	1.00					
26	0.27	0.12	0.22	0.44	0.88	1.00				
27	-0.11	-0.24	0.25	-0.09	-0.08	-0.04	1.00			
28	0.04	0.30	-0.40	0.16	0.05	0.03	-0.76	1.00		
29	-0.05	0.23	-0.44	0.12	00.05	-0.08	-0.65	0.77	1.00	
30	-0.04	0.09	-0.13	0.17	-0.04	-0.08	-0.49	0.53	0.56	1.00
31	-0.05	0.05	-0.41	-0.08	-0.06	-0.04	-0.25	0.36	0.34	0.14
32	0.64	0.38	0.32	-0.10	0,38	0.38	-0.11	0.02	-0.17	-0.14

	31	32
31	1.00	
32	-0.15	1.00

TABLE #3

Basic Regression Models

Variable Number	Regression Coefficients Model #1	Model #2	Model #3
1	1.69286*	-1.71993	2.14492
2	0.03976*	0.15018*	-0.14500*
5	-0.06395	1.98264*	
6			4.04246*
9			-0.67034
12	-0.25715	-0.44200	-1.60669
13	-0.12966	0.90354	2.52692
16	-0.24355	0.27384	
18	0.22122	0.25557	
20	0.00015*	0.00035*	0.00005
21	-0.04297	0.17903	-0.18160
23	0.54059	1.54168	-1.33990
26	0.01209	0.08689	
27	-0.00394	-0.00771	0.00440
28			0.01718
29	0.00992	0.02502	
30			0.15139
31			-0.12291
33			0.09672
Intercept	2.62303	74.72534	-33.80068
Multiple Correlation	.86973**	.66454**	.51381
N	128	128	128

* indicates that regression coefficient is significant at the .05 level.

** indicates that the regression is significant at the .01 level.

The intercorrelations matrix suggests that teachers found in the inner urban school identified by districts that service a population with a low median income tend to be black have fewer school-age children, and are less well academically prepared than their outer-city counterparts. The schools found in the inner-city tend to have a lower pupil/teacher ratio, have a higher percentage of black teachers, have less space per student, have more classrooms, and have had less recent improvements and renovation of school buildings than schools outside the inner city. Not surprisingly, parents of students in the inner-city tend to be less well educated, and their children's attendance rates and reading achievement scores tended to be somewhat lower than those found in outer-urban schools.[4]

Basic Regression Models

In the first two basic models the writers were interested in predicting student performance. In the first model the writers used 6th grade reading achievement (var 32) as a criterion measure. In the second model, attendance as a percent of enrollment is used as the dependent variable (var 22). Here the writers assumed that attendance rate provided a reasonable proxy measure of student attitudes toward schooling. In the third, and final model, the writers were interested in identifying the correlates of school holding power vis a vis its teaching staff. The average number of years of teaching experience at a particular school was used as a criterion measure (var 24).

All three basic regression models reported in Table 3 are significant at the .01 level. The coefficients indicate that reading achievement is significantly related to four independent variables. The positive coefficients associated with the percentage of white students at a particular school and median family income of parents underline the importance of the home factor in effecting student performance. Likewise, the sign of the coefficient associated with variable 23 (the ratio of capacity to enrollment) suggests that student overcrowding is associated with poor student academic performance. On the other hand, we would expect that the pupil/teacher ratio (var 2) would be negatively related to student performance. The result in Model 1 runs contrary to this expectation. Remembering, however, that our description of the inner city school showed that it tended to have lower pupil/teacher ratios at the particular point in time that data were collected suggests that these results might be expected. We might very well find that the impact of low pupil/teacher ratios might have the expected impact on student performance with the passage of time. This, of course, is something quite different than saying they would be enough to overcome the importance of home factors in effecting student performance.

Our second model, which uses attendance as a percent of enrollment as a criterion measure, also indicates the importance of home factors in determining student performance. The coefficients associated with median income (var 20) and education level of parents (var 21) are both significant and positively related to attendance rates. The positive sign associate with variable 5 (percent of teachers under forty years of age) suggests that students are more likely to attend classes taught by younger rather than older teachers. (Variable 2--pupil/teacher ratio) in Model 2, as in Model 1, shows a significant and positive relationship with student performance. Again, the only reasonable explanation the writers can offer is that the relationship resulted from changes that occurred in the district shortly before the data were gathered.

---

[4] In addition to the inspection of the intercorrelation matrix the writers also ran a series of three regressions using a binary coded median income criterion. The independent variables in each of these runs were teacher, school, and student variables as identified in Table 1. The results correspond to the results reported above.

The teacher retention equation indicates that teachers born in the area served by the district were most likely to stay with the district over periods of time. The model also shows that schools with high pupil/teacher ratios have a more difficult time holding teachers than schools where the reverse condition holds. Again, the reader is reminded of the behavior of this variable in the preceding performance equations. Nevertheless, it is interesting to note that low pupil/teacher ratios seem to effect the holding power of a school vis a vis its teachers, but do not effect student performance in the same way. Indeed, in the student performance models, the relationship is precisely the reverse.

SECTION II

Analysis of Data

The investigators employed two approaches in their analysis of the data. The first approach utilized the techniques of Ward[5] to determine the unique contribution of proper subsets of the predictor variables to three criteria. The unique contribution is defined to be as the difference between two squares of multiple correlation coefficients (R^2s), one obtained for a regression model in which all predictors are used, called the full model (FM), and the other obtained for a regression equation in which the proper subset of variables under consideration has been deleted; this model is called the restricted model, (RM). The difference between the two R^2s may be tested for statistical significance with the variance ratio test. The hypothesis tested states, in effect, that these variables contribute nothing to the determination of the expected criterion values that is not already available in the restricted prediction system.

The first model to be considered used as its criterion measure the sixth grade reading scores. Sixteen independent variables (1,2,5,12,13, 16,17,18,20,21,23,24,26,27,28,29) were used for the full regression model. In addition, these predictor variables were sub-grouped *a priori* into three disjoint subsets and the unique contribution of each of the subsets was tested for significance. Each of the three subsets was broken down further and the unique contribution of each component was tested at each stage. (Table 4 contains the various groupings and results of unique contribution tests.) The first subset (variables 1, 20, and 21), which might be called a home factor, had a significant unique contribution (see Table 4). Breaking the subset down further, variable 1 (percent white) and variable 20 (median family income) seemed to be making significant contributions to the explanation of the criterion of reading achievement. The unique contribution of the second subset (variables 2, 23, 27, 28, and 29) was significant beyond the .05 level. This particular subset might be considered a physical facilities factor. The ratio of capacity to enrollment (variable 23) emerged with the highest significant unique contribution as the analysis was extended. Finally, the third subset of predictor variables (variables 5, 12, 13, 16, 17, 18, 24, and 26), which might be considered as a teacher characteristics factor, failed to make a significant unique contribution to the explanation of the dependent variable.

Changing the criterion variable from reading achievement to attendance as a percent of enrollment (variable 22) and retaining the same sixteen predictors, the investigators found that the first subset again made a significant unique contribution (see Table 5). The principal contribution came from variable 20 (median family income). The physical facilities factor, the second subset, made a significant contribution with variable 28 (date of latest addition), variable 2 (pupil/teacher ratio), and

---

[5]Ward, J.H., "Multiple Linear Regression Models," Computer Applications in the Behavioral Sciences, Harold Borko (Editor), Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1962, pp. 204-237.

variable 23 (ratio of capacity to enrollment) appearing as important contributors. The teacher characteristics factor subset failed again to make a significant contribution. However, it is interesting to note that variable 5, which is contained in this subset did make a significant contribution on its own merit even though the total subset fell short.

(Table 4 contains the various groupings and results)

TABLE 4

Proportions of Variance Attributable to Groups of Variables Believed to be Associated with Sixth Grade Reading Scores

PREDICTOR-Variable Group	Total Contribution Proportion (R^2)	PREDICTOR-Variable Group	Unique Contribution Proportion
Model 1 (1,2,5,12,13,16, 17,18,20,21,23,24,26, 27,28,29) - Full Model (FM)	.7587		
Model 2 (FM - 1,20,21)	.6576	Variables 1,20,21	.1011[a]
Model 3 (FM - 1)	.7237	Variable 1	.0350[a]
Model 4 (FM - 20)	.7320	Variable 20	.0267[a]
Model 5 (FM - 21)	.7563	Variable 21	.0024
Model 6 (FM - 2,23,27,28,29)	.7260	Variables 2,23, 27,28,29	.0327[b]
Model 7 (FM - 27,28)	.7532	Variables 27,28	.0055
Model 8 (FM - 27)	.7550	Variable 27	.0037
Model 9 (FM - 28)	.7586	Variable 28	.0000
Model 10 (FM - 2)	.7511	Variable 2	.0076
Model 11 (FM - 23)	.7491	Variable 23	.0096[b]
Model 12 (FM - 29)	.7570	Variable 29	.0017
Model 13 (FM - 5,12,13,16 17,18,24,26)	.7455	Variables 5,12,13 16,17,18,24,26	.0132
Model 14 (FM - 16,24,26)	.7540	Variables 16,24,26	.0047
Model 15 (FM - 16)	.7574	Variable 16	.0013
Model 16 (FM - 24)	.7566	Variable 24	.0021
Model 17 (FM - 26)	.7570	Variable 26	.0017
Model 18 (FM - 17,18)	.7577	Variables 17,18	.0010
Model 19 (FM - 17)	.7586	Variable 17	.0001
Model 20 (FM - 18)	.7579	Variable 18	.0008
Model 21 (FM - 5,12,13)	.7570	Variables 5,12,13	.0017
Model 22 (FM - 5)	.7587	Variable 5	.0000
Model 23 (FM - 12)	.7572	Variable 12	.0015
Model 24 (FM 13)	.7587	Variable 13	.0000

[a] These proportions reported as unique contributions are significant at the .01 level for N = 128. In computing F values, it was assumed that one parameter was associated with each variable in the prediction system. The degrees of freedom for the number of predictors were determined by the number of variables given an opportunity to contribute to the prediction.

[b] Significant at the .05 level.

TABLE #5

Proportions of Variance Attributable to Groups of Variables
Believed to be Associated with Attendance as a Percent of Enrollment

PREDICTOR-Variable Group	Total Contribution Proportion (R^2)	PREDICTOR-Variable Group	Unique Contribution Proportion
Model 1 (1,2,5,12,13,16, 17,18,20,21,23,24,26, 27,28,29) - Full Model (FM)	.4624		
Model 2 (FM-1,20,21)	.2984	Variables 1,20,21	.1640[a]
Model 3 (FM-1)	.4508	Variable 1	.0116
Model 4 (FM-20)	.4049	Variable 20	.0575[a]
Model 5 (FM-21)	.4473	Variable 21	.0152
Model 6 (FM-2,23,27,28,29)	.3361	Variables 2,23,27,28,29	.1263[a]
Model 7 (FM-27,28)	.4330	Variables 27,28	.0294
Model 8 (FM-27)	.4622	Variable 27	.0003
Model 9 (FM-28)	.4417	Variable 28	.0207[b]
Model 10 (FM-2)	.4090	Variable 2	.0534[a]
Model 11 (FM-23)	.4206	Variable 23	.0418[a]
Model 12 (FM-29)	.4624	Variable 29	.0001
Model 13 (FM-5,12,13,16,17, 18,24,26)	.4199	Variable 5,12,13,16,17, 18,24,26	.0425
Model 14 (FM-16,24,26)	.4439	Variables 16,24,26	.0185
Model 15 (FM-16)	.4617	Variable 16	.0007
Model 16 (FM-24)	.4617	Variable 24	.0007
Model 17 (FM-26)	.4449	Variable 26	.0175
Model 18 (FM-17,18)	.4616	Variables 17,18	.0008
Model 19 (FM-17)	.4624	Variable 17	.0000
Model 20 (FM-18)	.4617	Variable 18	.0007
Model 21 (FM-5,12,13)	.4268	Variables 5,12,13	.0356
Model 22 (FM-5)	.4360	Variable 5	.0264[b]
Model 23 (FM-12)	.4610	Variable 12	.0015
Model 24 (FM-13)	.4557	Variable 13	.0067

[a] These proprotions reported as unique contributions are significant at the .01 level for N = 128. In computing F values, it was assumed that one parameter was associated with each variable in the prediction system. The degrees of freedom for the number of predictors were determined by the number of variables given an opportunity to contribute to the prediction.

[b] Significant at the .05 level.

TABLE #6

Proportions of Variance Attributable to Groups of Variables Believed to be Associated with Years Experience at Present School

PREDICTOR-Variable Group	Total Contribution Proportion (R^2)	PREDICTOR-Variable Group	Unique Contribution Proportion
Model 1 (1,2,5,6,12,13,17,18, 20,22,23,27,28,32) - FM (Full Model)	.3064		
Model 2 (FM-1,20,22,32)	.2819	Variables 1,20,22,32	.0245
Model 3 (FM-1,20)	.2829	Variables 1,20	.0235
Model 4 (FM-20)	.3011	Variable 20	.0053
Model 5 (FM-1)	.2838	Variable 1	.0226
Model 6 (FM-22,32)	.3025	Variables 22,32	.0040
Model 7 (FM-22)	.3057	Variable 22	.0007
Model 8 (FM-32)	.3025	Variable 32	.0039
Model 9 (FM-23,27,28)	.2551	Variables 23,27,28	.0513[b]
Model 10 (FM-27,28)	.2868	Variables 27,28	.0196
Model 11 (FM-27)	.3062	Variable 27	.0002
Model 12 (FM-28)	.2949	Variable 28	.0115
Model 13 (FM-23)	.2902	Variable 23	.0162
Model 14 (FM-2,5,6,12 13,17,18)	.1055	Variables 2,5,6,12, 13,17,18	.2009[a]
Model 15 (FM-5,6,12,13)	.1114	Variables 5,6,12,13	.1950[a]
Model 16 (FM-5)	.2334	Variable 5	.0730[a]
Model 17 (FM-6)	.2428	Variable 6	.0636[a]
Model 18 (FM-12)	.2992	Variable 12	.0007
Model 19 (FM-13)	.2663	Variable 13	.0072
Model 20 (FM-17,18)	.2994	Variables 17,18	.0401[b]
Model 21 (FM-17)	.2998	Variable 17	.0066
Model 22 (FM-18)	.3064	Variable 18	.0000
Model 23 (FM-2)	.2820	Variable 2	.0244[b]

[a] These proportions reported as unique contributions are significant at the .01 level for N=128. In computing F values, it was assumed that one parameter was associated with each variable in the prediction system. The degrees of freedom for the number of predictors were determined by the number of variables given an opportunity to contribute to the prediction.

[b] Significant at the .05 level.

TABLE No. 7

Principal Component Analysis of Sixteen
Predictors Used in Table 4 and Table 5

Variable	F_1	F_2	F_3	F_4	F_5
1	0.84	-0.01	-0.35	-0.06	0.14
2	-0.41	0.42	-0.45	0.53	-0.10
5	-0.66	-0.32	-0.25	-0.20	-0.29
12	-0.23	-0.12	0.14	-0.55	0.32
13	-0.79	-0.06	0.47	-0.00	-0.08
16	-0.49	-0.22	-0.58	-0.14	0.23
17	-0.41	-0.53	-0.21	-0.07	0.32
18	0.18	0.16	0.24	-0.36	-0.76
20	0.84	0.13	-0.36	0.01	0.01
21	0.70	0.24	-0.40	-0.10	-0.10
23	0.55	-0.47	0.16	-0.46	0.06
24	0.02	0.28	0.67	0.15	0.30
26	0.62	0.34	0.52	0.12	0.19
27	0.15	-0.76	0.15	0.40	-0.15
28	-0.27	0.83	-0.12	-0.24	0.13
29	-0.47	0.74	-0.08	-0.18	0.02
Eigenvalue	4.63	2.69	2.14	1.28	1.13
Cumulative Proportion of Total Variance	.29	.49	.61	.69	.76

TABLE NO. 8

Principal Component Analysis of
Fourteen Predictor Variables Used in Table 6

Variable	F_1	F_2	F_3	F_4	F_5
1	0.90	-0.09	0.19	0.12	-0.04
2	-0.28	0.63	0.49	-0.33	-0.02
5	-0.60	-0.10	0.14	0.01	0.58
6	0.29	0.08	-0.44	-0.45	0.14
12	-0.26	-0.14	-0.26	0.60	-0.11
13	-0.87	-0.02	-0.25	-0.06	0.07
17	-0.38	-0.39	0.49	0.27	0.38
18	0.17	0.08	-0.66	-0.22	0.34
20	0.90	0.09	0.17	0.03	0.05
21	0.31	0.49	0.10	0.00	0.60
23	0.47	-0.62	-0.24	0.26	0.27
27	0.00	-0.75	0.21	-0.46	0.00
28	-0.11	0.83	-0.13	0.34	0.02
32	0.89	0.15	0.12	0.07	0.12
Eigenvalue	4.26	2.51	1.47	1.21	1.08
Cumulative Proportion of Total Variance	.30	.48	.59	.68	.75

The third criterion variable investigated was variable 24 (See Table 6), years experience at present school. The fourteen predictors specified for this full model included variables 1,2,5,6,12,13,17,18,20, 22,23,27,28, and 32. The first subset consisted of variables 1,20,22 and 32. This particular subset of home factor variables did not make a significant unique contribution. The second subset consisting of physical facilities variables (23,27 and 28) made a significant unique contribution at the .05 level. None of the specific variables of this subset had a significant unique impact on the criterion, however. This might be explained by the high inter-correlations of these variables. Finally, the teacher factor subset (variables 5,6,12,13,17 and 18) was found to be making a significant (.01 level) unique contribution to the explanation of the criterion variable. A study of Table 6 reveals that variable 5 (percent under 40 years of age), variable 6 (percent raised in D.C.), variables 17 and 18 together (percent with bachelor's degree and percent with master's degree) and variable 2 (pupil/teacher ratio) were significant contributors to this subset.

In addition to the regression analysis with emphasis on unique contributions, the researchers sought to determine the unique contribution of factors to the explanation of the three criteria. Each set of predictor variables in the three regression models was factor analyzed using principal components and three new full regression models were generated in which each dependent variable was expressed as a function of the obtained factors.[6] In Table 7, the factors used for the first two regression runs are found. While there are 16 factors, only five were judged to be relevant.

Kaiser suggests that the number of factors judged significant be limited to those factors whose eigenvalues are greater than unity.[7] These five factors together account for 76 percent of the total variance of the sixteen independent variables; each of the remaining eleven factors contributes little to the over-all variance.

Using variable 32 as the criterion, a new regression model was investigated in which the five factors were utilized as independent variables. The unique contribution of factor 1 which loads heavily on variables 1, 13 and 20 (see Table 7) made a unique contribution which is estimated to be .5623. This was significant beyond the .01 level. The unique contribution of factor 2, estimated to be .0343, was also significant at the .01 level. This factor had high loadings on variable 27, 28 and 29. The estimated unique contribution of factor 3 (high loadings on variables 16 and 24) was .0924 which was significant unique contribution as the estimates in both cases are below .01. It is interesting to note that factor 1 is related to the home factor in the previous regression runs, while factor 2 seems related to the physical facilities and factor 3 emphasizes the teacher characteristics.

---

[6] For a detailed discussion of the process of determining the regression models, see: W. F. Massy, "Principal Components Regression in Exploratory Statistical Research," Journal of the American Statistical Association, March 1965, pp. 234-256.

[7] See W. W. Cooley and P. R. Lohnes, Multivariate Procedures for the Behavioral Sciences, Wiley, N. Y., 1962, p. 162.

In the second regression run, variable 22 served as the dependent variable. When the five factors used with criterion variable 32 were used as predictors of variable 22 (attendance as a percent of enrollment), the same three factors emerged as significant. Factors 2 and 3 were significant at the .01 level while factor 1 was significant at the .05 level. Factor 2 appeared to be the dominant contributor with its unique contribution estimated to be .1178.

Using variable 24 as a criterion, a different set of 14 independent variables served as predictors. When these 14 variables were factor analyzed, five factors were identified to be relevant using Kaiser's rule for significant contribution. These five factors appear in Table 8, and together they account for 75 percent of the total variance of the fourteen independent variables; the other 25 percent is distributed over the remaining nine factors. Of the five factors, only factor 3 made a significant unique contribution to the explanation of the criterion variable 24. Its contribution was estimated to be .0748, which was significant beyond the .01 level. The high loadings appear to be on variables 2, 17, and 18. These variables provide information about the teacher.

It was hoped that the unique contribution approach and the factor-regression models would supply information which might be complementary. The results of both approached suggest that they are indeed comparable. This can be explained by the fact that the a priori specification of the three subsets to be analyzed turned out in reality to be related to the factors obtained in the principal components analysis.

Conclusions:

The results of this study indicate that home factors, specifically the median income and the education level of parents are more important than school or teacher considerations in influencing student performance. The study also suggests that the vitality of a youthful teaching staff is important in improving student activities toward schooling.

With regard to the retention of teachers our results suggest that school and teacher factors are more important than student considerations. Teachers raised in or near the district of their employment are more likely to remain with that district over periods of time than teachers recruited from other areas. Also the results indicate that while a low pupil/teacher ratio seems to improve the ability of the district to hold teachers it does not appear to improve student performance. For reasons stated previously, however, our conclusion with regard to pupil/teacher ratios must be considered high tentative.

1971 MEMBERSHIP LIST
Bill Connett, Secretary

Auton, Sylvia
c/o Department of Education
Measurements and Statistics
University of Maryland

Ashburn, Arnold G.
Deputy Assistant Superintendent
Research and Evaluation
Dallas Independent School District
3700 Ross Avenue
Dallas, Texas 75204

Bill Weber
Research and Evaluation
Dallas Independent School District
3700 Ross Avenue
Dallas, Texas 75204

Barbuto, Paul, Jr.
Box 37
Teachers College
Columbia University
New York, New York 10027

Connett, William E.
Dept. of Research and
Statistical Methodology
University of Northern Colorado
Greeley, Colorado 80631

Eichelberger, Tony
Coordinator-Research Design
Dallas Independent School District
3700 Rose Avenue
Dallas, Texas 75221

Ernhart, Claire B.
10670 St. Chas. Rock Rd.
St. Ann, Missouri 63074

Foster, Garrett R.
Assistant Professor
405 Education
Florida State University
Tallahassee, Florida 32306

Green, Beatrice
East Street
Stockbridge, Massachusetts

Halasa, Ofelia
Division of Research and Development
Cleveland Public Schools
Cleveland, Ohio

Harrison, Joseph E.
13200 St. Louis Avenue
East St. Louis, Illinois 62201

Hendrix, Vernon L.
Professor
221 Borton Hall
University of Minnesota
Minneapolis, Minnesota 55455

Hick, Thomas L.
Director, Child Study Center
Campus School
State University College
New Paltz, New York 12561

Houston, Samuel R., Associate Professor
Dept. of Research and Statistical Meth.
University of Northern Colorado
Greeley, Colorado 80631

Hyde, Janet Carol
1048 M Graduate House
West Lafayette, Indiana 47906

Jordon, Thomas E.
EDAP Program Director
CEMREL
10646 St. Charles Rock Road
St. Ann, Missouri 63074

Kelly, Francis J.
Professor of Educational Psychology
Educational Research Bureau
Southern Illinois University
Carbondale, Illinois 62901

V. Rutledge McClaran
1722 W. Oak
Denton, Texas 76201

McNeil, Keith A.
Dept. of Guidance and Educ. Psychology
Southern Illinois University
Carbondale, Illinois 62901

Martin, R. A.
College of Education
University of Toronto
371 Bloor St. W
Toronto, Canada

Merrill, Paul F.
Florida State University
1A Tully Building
Tallahassee, Florida 32306

Newman, Isadore
Park Towne Garden Apts.
Carbondale, Illinois 62901

Powell, Marjorie (Miss)
20 Dartmouth Place
Boston, Massachusetts 02116

Reed, Cheryl L.
Rt. 9, Box 72
W. Lafayette, Indiana 47906

Ritter, Carolyn E.
Computer and Data Processing Center
Carter Hall
University of Northern Colorado
Greeley, Colorado 80631

Ritter, Emmett A., Associate Professor
Educational Administration
McKee 419
University of Northern Colorado
Greeley, Colorado 80631

Russell, Nolan F.
4507 Berkley Street
Harrisburg, Pennsylvania 17109

Spaner, Steven D.
School of Education, UMSL
8001 Natural Bridge Road
St. Louis, Missouri 63121

Thomas, Donald L.
76 Loretta Avenue, Apt. L.
Fairborn, Ohio 45324

Ward, Joe H., Jr.
Southwest Educational Laboratory
167 East Arrowhead Drive
San Antonio, Texas 78228

Wasen, G. Leighton
50 Wedgewood Terrace
Springfield, Illinois 62702

Williams, John D.
Bureau of Educational Research
University of North Dakota
Grand Forks, North Dakota 58201