DOCUMENT RESUME

ED 076 691 TM 002 700

AUTHOR Marco, Gary L.

TITLE An Example of the Application of the Assessment and Diagnostic Procedures of a Comprehensive Accountability Plan.

PUB DATE 73

NOTE 21p.; Paper presented at annual meeting of American Educational Research Association (New Orleans, Louisiana, February 25-March 1, 1973)

EDRS PRICE MF-$0.65 HC-$3.29

DESCRIPTORS *Educational Accountability; *Educational Diagnosis; Grade 3; *Performance Factors; School Improvement; Standardized Tests; *Student Evaluation; Techr' 'al Reports

ABSTRACT

The assessment and diagnostic procedures of a comprehensive accountability plan were applied to several elementary schools from a large midwestern state. Pretest and posttest Word Knowledge and Reading scores from the Primary II Metropolitan Achievement Test administered in 1970-71 to third-graders were used. These data were used to compute Student Development Indices (SDIs). The SDIs were related to school process variables. Application of the diagnostic procedures resulted in the identification of the total and unique contributions of the process variables, individually and in combination. The implications of the assessment and diagnosis for corrective action are discussed. (Author)

AN EXAMPLE OF THE APPLICATION OF THE ASSESSMENT AND DIAGNOSTIC PROCEDURES OF A COMPREHENSIVE ACCOUNTABILITY PLAN

by

Gary L. Marco

Educational Testing Service

Paper presented at the meeting
of the American Educational
Research Association, New Orleans,
February - March, 1973

AN EXAMPLE OF THE APPLICATION OF THE ASSESSMENT AND DIAGNOSTIC PROCEDURES OF A COMPREHENSIVE ACCOUNTABILITY PLAN[1]

Gary L. Marco

Educational Testing Service

A few days ago I took a boat ride into bayou country and learned the difference between a marsh and a swamp. It seems that a marsh is different from a swamp in the same way that a prairie is different from a forest. If I asked a native what the difference between a marsh and a swamp was, I'm sure he wouldn't think of this analogy. He'd probably say something like, "Well, son, if you go out wandering and get lost, you'll know you're in a swamp."

Today I shall take you through the marshland of an accountability plan. We'll look at a few of the prominent features having to do with assessment and diagnosis. I hope that the marshland doesn't turn into a swamp before we finish. To point you on your way I want first to give you a brief introduction to the accountability plan I have in mind. After that we'll look at an example that shows how the assessment and diagnostic procedures of the plan can be applied.

[1] Presented at the meeting of the American Educational Research Association, New Orleans, February - March 1973.

The Accountability Plan

The accountability plan was designed for elementary and secondary schools, and focuses on the school as the basic unit of accountability. It includes many features of Dyer's student-change model, which was described in the Phi Delta Kappan (Dyer 19[illegible]).

The plan consists of three phases: assessment, diagnosis, and corrective action. In the assessment phase the achievement of students is measured at two or more points in time. Data based on students who have stayed in the same schools over the last two data collection points are used to derive measures of school effectiveness. I should point out that such longitudinal samples may be small in schools having a highly transient student population. (It is interesting that schools often do not have an accurate indices of student stability, although they do have indices of student mobility. Since the same students who transfer in often transfer out, a student mobility index can give a very misleading estimate of the number of students who have remained in the school over a given period of time.) The accountability plan assumes that assessment data are available over a two-year period. However, the plan can be adapted to longer or shorter periods of time. And, in fact, the illustrative data that I will use covers only a one-year period.

Key to assessment is the computation of one or more indices of student development for each school. The index is a number that expresses the extent to which students at a given grade level in the school have progressed in a particular area of achievement over a specified length of time. The index takes into consideration where a student was and where student stands now, but as you will see it is not the usual gain measure.

The student development indices serve to identify schools in which students are performing relatively poorly. It is these schools that need to take corrective action. However, before corrective action is taken, the school staffs need some guidelines to go on. It is the purpose of the diagnostic phase of the accountability plan to provide these guidelines by identifying possible causal variables that might be working at the school level.

In the diagnostic phase the various variables that might be related to student development must be sorted into those that corrective-action plans might focus on and those that the plans can ignore. These variables are called process variables. They refer to any characteristic of the school program, facilities, students, staff, or community which might reasonably be expected to relate to student development and over which the school can exercise some degree of control. Diagnosis proceeds in two directions: (1) the statistical analysis of the available data to show the relative contributions of the various process variables to student performance; (2) case studies of schools having high and low indices of student development. The case studies are particularly useful for uncovering important process variables that were unmeasured to begin with.

In the last phase of accountability, corrective action, the schools having low indices of student development are required to produce correction-action plans. These plans would ordinarily focus on variables that were identified in the diagnostic phase as important. It is the diagnostic and corrective-action phases that make the accountability plan different from the usual school assessment plan, which stops with an evaluation of student performance. Accountability is thus defined in terms of the responsibility of school personnel in designing, implementing, and evaluating corrective action.

So far, I have given you a brief but, I hope, adequate description of the essential components of a comprehensive accountability plan. In the time remaining I shall concentrate on procedures used to compute the indices of student development and procedures used to assess the importance of process variables.

Sample

The procedures will be applied to Metropolitan Primary II Reading Achievement Test data from six elementary schools. The test was administered to third-graders in the fall and spring of the 1970-71 academic year. For demonstration purposes the Word Knowledge and Reading scores will be used. Pretest and posttest means for those students who took tests in both the fall and the spring are shown in Table 1. Note that even though Schools D and E had similar pretest means, School E had

Insert Table 1 about here

much higher posttest means. Moreover, School F had lower pretest means than School C but a higher Reading posttest mean. We would expect the indices of student development to reflect the superior performances of Schools E and F.

Assessment

The student development indices for the schools are really mean individual residual scores and are not particularly difficult to compute. First the regression coefficients were estimated for predicting the Reading posttest and Word Knowledge posttest scores, respectively, from both Reading and Word Knowledge pretest scores. The scores used were those of all 238 third-graders having pretest and posttest scores.

The models used here did not employ higher-order terms such as the squares of the pretest scores, but there is no reason why such terms could not be added if the relationships were not assumed to be linear. In actual practice, the coefficients for a complete second-order model, employing first- and second-degree terms, would probably be estimated and tested for statistical significance. Cross-validation would also be used to ensure that the coefficients would work on new samples.

The coefficients for the two regressions were:

$$\hat{R}_{post} = 13.38927 + 0.46532\ WK_{pre} + 0.39521\ R_{pre}, \text{ and}$$

$$\hat{WK}_{post} = 13.74727 + 0.52659\ WK_{pre} + 0.35294\ R_{pre}.$$

The mean individual residual is simple to calculate once the regression coefficients have been estimated. The expected Reading score ($\hat{R}_{post}$) for a school is computed by substituting the Word Knowledge and Reading pretest means for the school into the first regression equation. Thus, the expected Reading score for School A is 13.38927 + 0.46532 (46.2308) + .39521 (41.0256), which equals 51.1151. This value is the expected Reading posttest performance based on the average performance of students in the six schools. The mean individual residual is equal to the posttest mean for the school minus the expected mean. In the case of School A, this is 46.8718 - 51.1151 = -4.2433. The number is negative, which indicates that the performance of the average student in this school in reading was below the average for all students.

The 12 mean individual residuals that were computed along with the posttest means and expected performances are shown in Table 2 for both

---

Insert Table 2 about here

Reading and Word Knowledge. These mean individual residuals are displayed graphically in Figures 1 and 2, which are in your handout. You may note that the superiority of Schools E and F is quite apparent. On the other

---

Insert Figures 1 and 2 about here

---

hand, Schools A and D are below the line of prediction and would be singled out for special attention. (Of course, it would be useful to establish confidence limits for the predicted values to ensure that the deviation from prediction did not arise from sampling fluctuation.)

Before turning to procedures used for diagnosis, I should point out that my discussion of the assessment phase is incomplete. Quirk (1973) presents a more complete discussion.

Diagnosis

It is not enough to know that the students in some schools are doing well and others not so well in Reading and Word Knowledge. Something must be done to improve performance. The diagnostic component of the accountability plan is designed to help staff know what has worked within the system. Process information from the various schools is assembled, reduced, and related to student performance. At this stage the relationships are only correlational, not causal. Later, when corrective-action plans are formulated, the assumption of causation must be made. Identifying important process variables is much like trying to find out the cause of your stomach-ache after having breakfast at Brennan's, cocktails at the Top of the Mart, and dinner at Antoine's.

Initial Data Analysis

In this small-scale tryout of the accountability procedures only five process variables were used: pupil/teacher ratio (P/T), percent of teachers with five or more years experience (%EXP), the percentage of non-white students (%N-W), the K-12 current operating expense per pupil ($/PUPIL), and the number of third-graders (N). You may argue that some of these are not under the control of the school and thus are not good choices for process variables. However, of these variables, most can be changed at the school district level if not the school level. If a large number of variables had been available, factor-analytic methods could have been used to reduce the data to the desired number of variables.

There are two common kinds of analyses that are less than informative than the one I propose. The first is a multiple regression analysis of individual pupil scores with the pretest scores and the process variables used as independent variables. This analysis ignores among-school variation.

The second is a regression analysis using (a) the school Mean Residuals obtained by partialing out the student pretest scores from the student posttest scores and (b) the process variables. This analysis utilizes only among-school variance and ignores within-school variance. The analysis I shall illustrate makes use of both among-school and within-school variation.

There is a third analysis that some might use. It is an analysis of the variance associated with the process variables and the among-school variance. This kind of analysis is wrong if these variables are treated simultaneously in a regression analysis. The approach I suggest looks at among-school variance and the variance associated with process variables sequentially. The key to this analysis is obtaining the among-school and within-school sums of squares before looking at the process variables.

The first step in the analysis is to eliminate the influence of the Reading and Word Knowledge pretests. This is done by treating the pretests as concomitant variables in a multiple regression analysis. If you will refer to Table 3 you can see that in the illustrative data the pretests accounted for more than half of the total sum of squares for Word Knowledge

---

Insert Table 3 about here

---

and Reading. The next step is to compute the sum of squares associated with differences among the school intercepts, assuming common within-group regression coefficients. Finally, the sum of squares accounted for by differences in the regression coefficients for the pretests is computed. The results of these analyses are also shown in Table 3. Note that the hypothesis of common regression coefficients is rejected at $\alpha = .05$ for Word Knowledge and for Reading.

If there were no differences among the regression coefficients or intercepts, it would not be worthwhile to search for school process variables that account for among-school variance. Rather one would have to look within the schools to identify important process variables associated with either classes or individual students.

Here the statistical tests permit us to continue. We proceed to look at the contributions of the five process variables. Table 4 shows how the schools stood on these variables. Computationally, the five

---

Insert Table 4 about here

---

process variables were simply substituted into the regression program in place of the school dummy variables. A forward-selection stepwise regression procedure was then used to order the process variables in terms of their contributions to the school intercept sum of squares. The results are given in Table 5. The error terms for the F-tests were the residual mean squares from Table 3.

Insert Table 5 about here

In the case of both Word Knowledge and Reading, the same three process variables accounted for practically all of the variance associated with differences in school intercepts. These were the percentage of non-white students, the number of third graders, and the percentage of teachers with five or more years of teaching experience.

We note in passing that if the degrees of freedom for schools had exceeded the number of process variables, it would have been possible to test the residual school intercept variance to see if adding additional process variables would have been very profitable. You may note that the variance associated with the school regression coefficients was not analyzed. In a full-scale analysis, the contribution of the process variables to this variance would also be investigated.

Newton-Spurrell Procedure

One final step remains to complete the analysis. We need a measure of the importance for each of the three process variables selected by stepwise regression. The proportion of sum of squares accounted for as each variable is added in the stepwise solution is not a good measure. Since the variables are correlated with each other, changing the order in which the variables are added changes the sum of squares accounted for. The procedure suggested for use in the accountability plan is the one described by Newton and Spurrell (1967) and used by Mayeske in reanalyzing the Coleman data (Mayeske et al., undated). This procedure partitions the variance associated with a given variable into the portion accounted for by the variable uniquely and the portions accounted for by the variable in combination with other variables. To do this partitioning, we must obtain the proportion of variance accounted for in the Reading and Word Knowledge posttest scores by each of the six possible combinations of the three process variables. These proportions are listed in Table 6 in the column labelled "Total Contribution." t contribution

Insert Table 6 about here

is simply the squared multiple correlation. The unique contributions were computed from the total contributions. If you will look at the Venn diagrams in Figure 3, you will note that the unique contribution

Insert Figure 3 about here

of a given variable is simply the total contribution of all three variables minus the total contribution of the other two variables. Thus, the unique contribution of % Non-White in Figure 3a is .0291. Once the unique contributions of the individual variables are known, it is relatively easy to find the other unique variances. For example, unique contributions of % Non-White and N combined is what is left of the <u>total</u> contribution after the total contribution of the percentage of teachers with five or more years of teaching experience and the unique contributions of the other two variables are taken out.

So much for the computations. If the marshland has turned into a swamp for some of you, maybe we can start finding our way out now. We wish to select the variables we might like to focus on in a corrective-action plan. We should probably concentrate on those combinations that have (a) high unique contributions relative to their total contributions and (b) high total contributions. Where do these rules lead us in the present case? We note in Table 6 that for Word Knowledge % Non-White had a reasonably high total contribution and that its unique contribution was nearly half of its total contribution. So % Non-White seems to be an important variable in the case of Word Knowledge. For Reading the unique contribution of % Non-White is nearly equal to its total contribution, but its total contribution made up only about half of the total contribution of all three variables. However, the unique contributions of the combinations that have larger total contributions than % Non-White are so low that % Non-White seems to be the choice here also.

What does it mean for a corrective-action plan to focus on % Non-White. We must first note whether the regression weight for % Non-White is positive or negative. Here % Non-White had negative weights for estimating both Word Knowledge and Reading posttest scores. However, it would be misleading to treat % Non-White as a causal variable, even if the school district could decrease the percentage of non-white students. It is the job of technical staff to decide <u>why</u> % Non-White was related to student development in Word Knowledge and Reading. At this point the diagnostic procedure becomes more of an art than a science. On the basis of what is known about the schools, it may be decided that the schools having high proportions of non-white students have poor physical facilities or that they have few innovative reading programs. Somehow real underlying causes must be identified. The corrective-action plan would then focus upon the hypothesized causal variables. The important point here is that the diagnostic analysis suggested the direction in which to look for a causal influence but did not indicate necessarily what the causal influence was.

Once possible underlying causal variables are identified, the diagnostic phase is finished. Of course, case studies would be necessary if the process variables accounted for only a small portion of the among-school variance. In the accountability plan the results of the diagnosis would be written up in the form of guidelines. And, on the basis of the guidelines, corrective-action plans would be formulated. A school's corrective-action plan would focus on the causal variables thought to be underlying the important process variables. It should be clear from this example that analysis carries one only so far in diagnosis. Judgment must take over when causation is inferred and the corrective-action plan is formulated.

We started out today walking through a marshland. I hope the features I have illustrated will help us make our way to dry ground.

References

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

Mayeske, G. W., Wisler, C. E., Beaton, A. E., Weinfeld, F. D., Cohen, W. M., Okada, T., Proshek, J. M. and Tabler, K. A. A Study of Our Nation's Schools. Washington, D. C.: U. S. Department of Health, Education, and Welfare, Office of Education, undated.

Newton, R. G., and Spurrell, D. J. A development of multiple regression for the analysis of routine data. Applied Statistics, 1967, 16, 51-64.

Quirk, T. J. Assessing student performance. Paper presented at the meeting of the American Educational Research Association, New Orleans, February - March, 1973.

Table 1

Pretest and Posttest School Means

	School A (N=39)	School B (N=56)	School C (N=29)	School D (N=52)	School E (N=28)	School F (N=34)
Wd. Know. Pre	46.2308	50.3214	62.7241	50.4615	50.2500	57.1765
Reading Pre	41.0256	47.5714	60.6552	45.3462	45.7857	57.2647
Wd. Know. Post	48.3590	58.9643	68.2759	53.0769	59.6071	67.9412
Reading Post	46.8718	57.2679	64.3103	51.3077	58.7143	68.8235

Table 2

Expected Performances and Residuals for Each School[a]

	School A (N=39)	School B (N=56)	School C (N=29)	School D (N=52)	School E (N=28)	School F (N=34)
Wd. Know. Post	48.3590	58.9643	68.2759	53.0769	59.6071	67.9412
Wd. Know. Expected	52.5715	57.0359	68.1848	56.3243	56.3680	64.0668
Wd. Know. Residual	-4.2125	1.9284	0.0911	-3.2474	3.2391	3.8744
Reading Post	46.8718	57.2679	64.3103	51.3077	58.7143	68.8235
Reading Expected	51.1151	55.6055	66.5476	54.7913	54.8666	62.6262
Reading Residual	-4.2433	1.6624	-2.2373	-3.4836	3.8477	6.1973

[a]The expected performance was computed by substituting a school's pretest scores into the regression equation for all schools. The residual is the difference between the posttest mean and the expected performance.

Table 3

Analysis of Variance of Posttest Scores

Source of Variation	S.S.	df	M.S.	F	Pr
	WORD KNOWLEDGE				
Total (Centered)	32158.4244	237			
Pretests	19554.6634	2	9777.3317	231.712	<.01
School Intercepts	2500.3699	5	500.0740	11.851	<.01
School Regression Coefficients	820.2808	10	82.0281	1.944	<.05
Residual	9283.1103	220	42.1960		
	READING				
Total (Centered)	34547.1765	237			
Pretests	19228.4306	2	9614.2153	196.603	<.01
School Intercepts	3577.9198	5	715.5840	14.633	<.01
School Regression Coefficients	982.4469	10	98.2447	2.009	<.05
Residual	10758.3792	220	48.9017		

Table 4

Standings of Schools on Process Variables

School	$/PUPIL	% N-W	P/T	% 5+ EXP.	N
A	857	74	29.0	52	93
B	676	1	30.6	67	66
C	644	2	22.8	80	32
D	714	99	21.7	52	66
E	878	26	20.9	60	36
F	893	1	24.7	57	38

Table 5

Selection of Process Variables by Step-Wise Regression (Forward Method)

Source of Variation	S.S.	df	M.S.	F	Pr
	WORD KNOWLEDGE				
School Intercepts	2500.3699	5			
% N-W	1901.4511	1	1901.4511	45.062	<.01
N | N-W	336.4170	1	336.4170	7.973	<.01
% 5+ EXP. | N-W, N	200.9094	1	200.9094	4.761	<.05
$ / PUPIL | N-W, N, EXP.	44.7393	1	44.7393	1.060	>.10
P/T | N-W, N, EXP., $	16.8531	1	16.8531	0.399	>.10
	READING				
School Intercepts	3577.9198	5			
% N-W	2039.5025	1	2039.5025	41.706	<.01
% 5+ EXP. | N-W	919.2727	1	919.2727	18.798	<.01
N | N-W, EXP.	549.9113	1	549.9113	11.245	<.01
$ / PUPIL | N-W, N, EXP.	58.4553	1	58.4553	1.195	>.10
P/T | N-W, N, EXP., $	10.7780	1	10.7780	0.220	>.10

Table 6

Total and Unique Contributions of Various Combinations of Three Process Variables

	Word Knowledge		Reading	
Variable	Total Contribution	Unique Contribution	Total Contribution	Unique Contribution
% N-W	.0591	.0291	.0590	.0542
N	.0454	.0125	.0458	.0159
% 5+ EXP.	.0166	.0062	.0036	.0318
% N-W, N	.0696	.0176	.0697	.0278
% N-W, % 5+ EXP.	.0633	-.0049	.0856	-.0052
N, % 5+ EXP.	.0467	-.0020	.0473	-.0303
% N-W, N, % 5+ EXP.	.0758	.0173	.1015	.0073

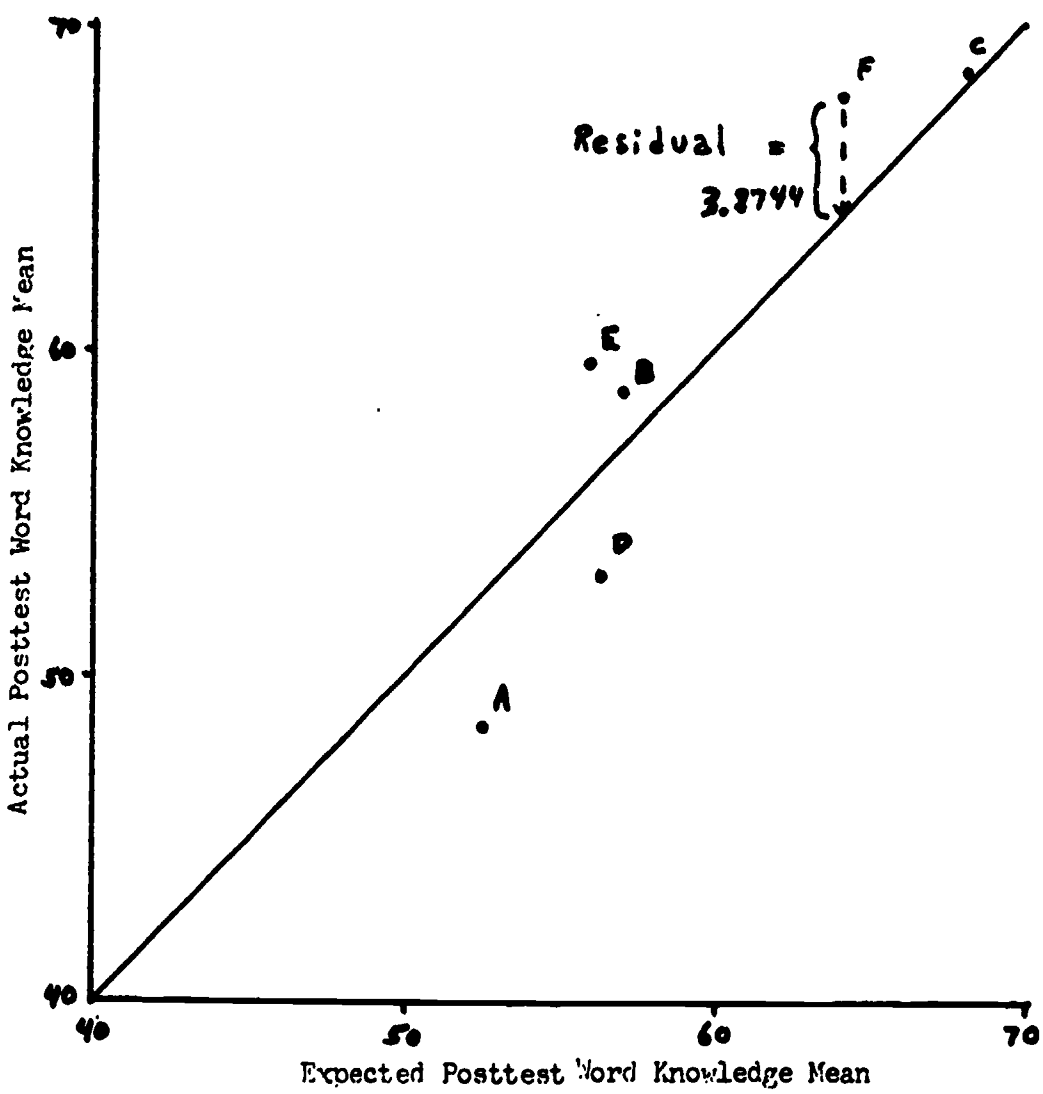


Fig. 1. Posttest Word Knowledge means plotted against expected performance.

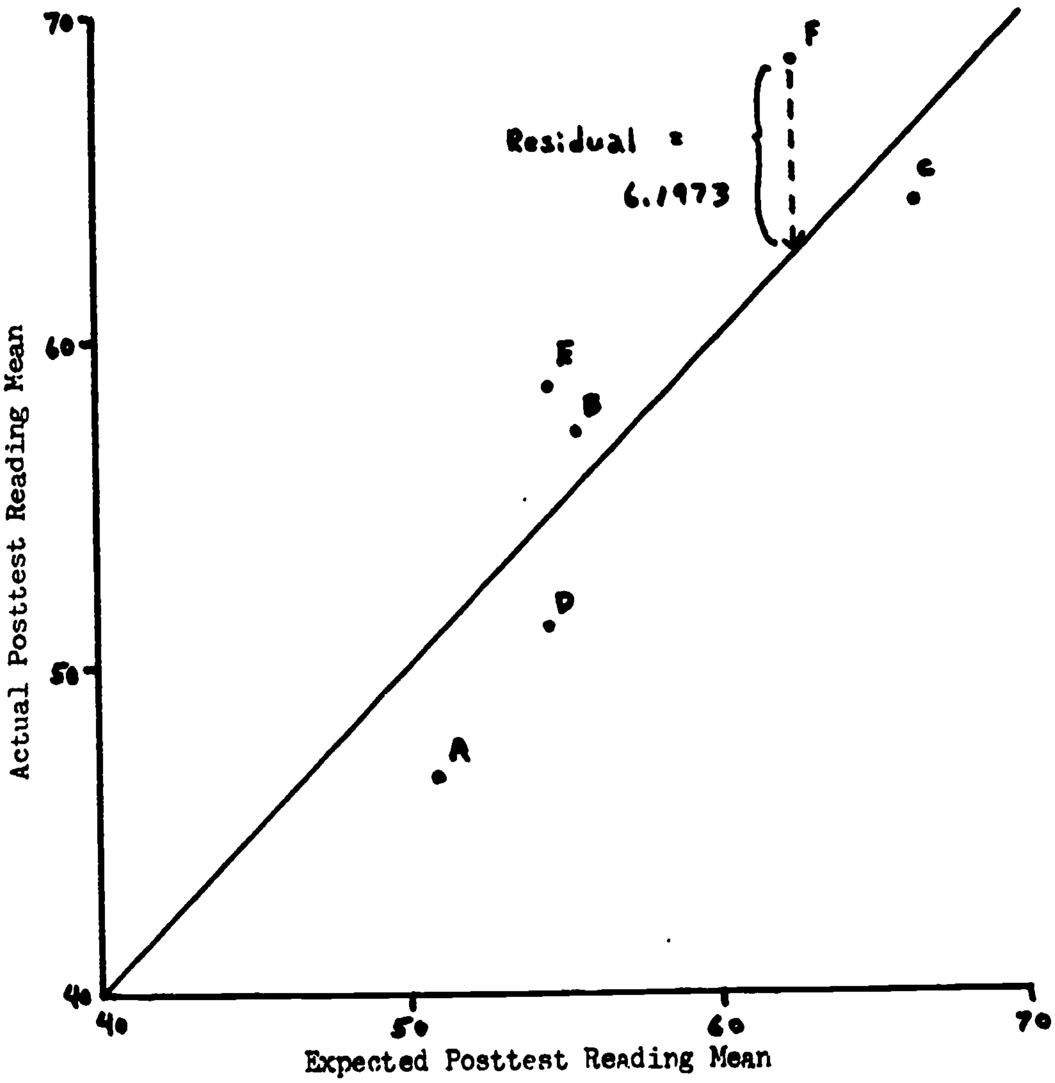
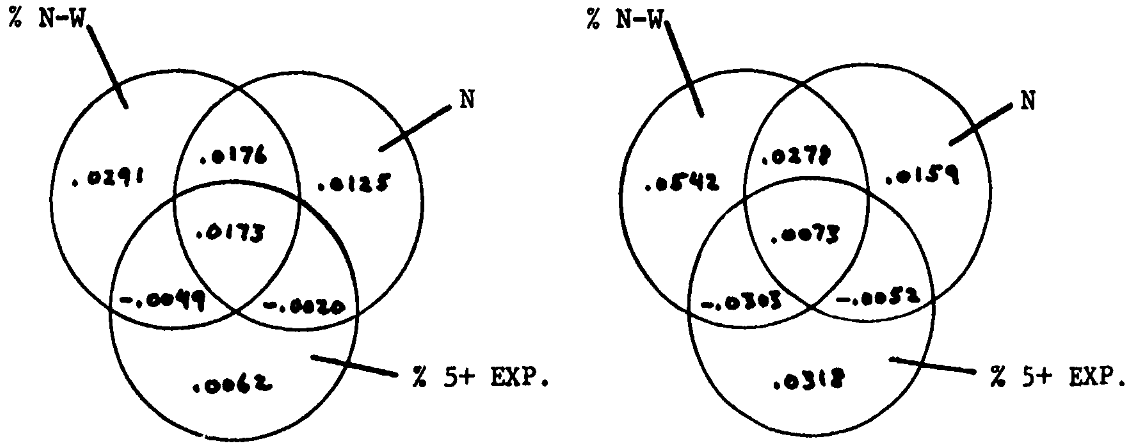


Fig. 2. Posttest Reading means plotted against expected performance.



(a) Word Knowledge (b) Reading

Fig. 3. Unique contributions of combinations of three variables to the variances of Word Knowledge (a) and Reading (b) posttest scores.