DOCUMENT RESUME

ED 191 912 TM 800 536

AUTHOR Deck, Dennis D.
TITLE Student Selection and the Special Regression Model.
PUB DATE Apr 80
NOTE 27p.: Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Rating: Achievement Tests: *Admission Criteria: *Compensatory Education: Elementary Education: Elementary School Mathematics: *Mathematical Models: *Multiple Regression Analysis: *Predictor Variables: Pretests Posttests: Quasiexperimental Design: Reading Achievement: Scores: Student Evaluation
IDENTIFIERS *Composite Scores: Elementary Secondary Education Act Title I: *RMC Models: Stanford Achievement Tests

ABSTRACT

The feasibility of constructing composite scores which will yield pretest measures having all the properties required by the special regression model is explored as an alternative to the single pretest score usually used in student selection for Elementary Secondary Education Act Title I compensatory education programs. Reading data, including Stanford Achievement Test scores, obtained from students in grades 2, 3, 4, and 6 in four school districts are analyzed from a technical and a practical perspective. Mathematics data obtained from students in grades 2 and 3 in one school district are also analyzed. Although composites do not seem to increase the accuracy of estimating the no-treatment expectation in regression designs variables such as teacher ratings and placement in a reading series can be quite reliable. A procedure requiring the simple sums of two or three carefully selected variables has obvious practical advantages over elaborate score transformations. It can yield pretest measures having all the properties necessary for the special regression model. If, however, the scaling of a variable or score inflation on a rating causes floor or ceiling effects in the composite scores, problems are created when project gains are estimated with the special regression model. (RL)

STUDENT SELECTION AND THE SPECIAL REGRESSION MODEL

Dennis D. Deck

Northwest Regional Educational Laboratory

A paper presented at the annual meeting of the American Educational Research Association, Boston, April 7-11, 1980.

The evaluation of educational programs in field settings where true randomization is not possible allows conditions to be associated with the treatment which may pose rival explanations of the results. Quasi-experimental designs, as Kenny (1979) points out, must have explicit models incorporating differences between the treatment and comparison groups over time. Since the selection of disadvantaged students for supplementary instruction is by nature non-random, the Title I evaluation system (Tallmadge & Wood, 1978) provides three alternative models for estimating the impact of compensatory education programs. This paper will focus on selection issues surrounding only one of these, the special regression model, sometimes referred to as Model C.

The special regression model may be used in situations where treatment students are identified solely on the basis of the pretest. The model assumes that the regression line computed from the comparison group pre- and posttest scores may be used to predict what the treatment group mean would have been without Title I instruction. In other words, the comparison group regression line is projected below a strict pretest cutoff to predict what the performance of the Title I group would have been without supplemental instruction. The difference between this "no treatment expectation" and the observed treatment group mean represents the gain in achievement attributable to Title I instruction.

The model assumes that pretest scores are linearly related to posttest scores and that the treatment and comparison group regression lines are parallel. These assumptions may not be met when there are floor or ceiling effects on either the pretest or posttest (Estes & Anderson, 1978). A moderately high correlation between the pretest and posttest and a reasonably large sample of students are recommended to ensure an accurate evaluation.

School districts have attempted to implement the special regression model with mixed success. The primary difficulty has been student selection. For practical or political reasons, many districts have been unable to maintain a strict pretest cutoff. Under these conditions the model is underspecified and will produce misleading results.

Typically the district evaluator will use one of the following procedures to analyze the data:

(1) Exclude treatment students above the cutoff and comparison students below the cutoff from the analysis.

(2) Include all students in the analysis, whether selected appropriately or not.

(3) Follow the analysis procedures for the norm referenced model (Model A).

All three analysis procedures are unsatisfactory. If the variable added to the selection process correlates with posttest scores, the first two procedures tend to underestimate project effectiveness. The first procedure may be used, however, when the number of students that must be eliminated is very small or when the added variable does not correlate with posttest scores, for example, if students must be reassigned due to random scheduling problems. The third procedure tends to overestimate project effectiveness due to regression to the mean since the pretest was used, in part, for selection.

A fourth procedure was proposed by Yap, Estes and Hansen (1979) who argued that an unbiased estimate of the treatment effect could be obtained by dropping all treatment and comparison students in a band where there was overlap between the two groups on the pretest. This procedure, however, will often result in a loss in precision since the wide band required in most situations would cause too many students to be eliminated from the analysis. The more restricted range of scores in these extreme groups would make the procedure more sensitive to floor and ceiling effects.

While there appear to be no acceptable *post hoc* procedures for applying the special regression model when students have not been assigned to the treatment solely on the basis of the pretest scores, more acceptable selection measures can be constructed. Several variables which the district wishes to use for student selection can be combined to form composite scores (Tallmadge, 1978). As long as a strict cutoff on this composite measure is maintained in selection, the composite scores may be used in the analysis. To construct the composite, the variables are either standardized or simply multiplied by constants so that each has the same standard deviation and thus the same contribution to the composite. A well constructed composite can increase the accuracy of the regression model and the validity of the selection procedure.

Despite the technical rationale for using composite scores, districts are skeptical. They want to know if all this extra work is worth it. Will composites improve the accuracy of Model C? Will the composite tend to select the students teachers think should be served? Can simple composites be designed that do not require excessive hand computation or expensive computer equipment? This paper reviews the empirical findings of previous attempts by districts to use composites and discusses the implications of these findings for Title I student selection and the special regression model.

FIELD APPLICATIONS OF COMPOSITE SCORES

Four applications of composite scores by school districts are reported here to explore issues in the implementation of composite scores for Title I student selection. The examples differ considerably in the measures selected for the composite, the way the composite was constructed, and the role the composite played in selection. This natural variation is useful for interpreting the results of each district. It should be noted that each of the districts designed and carried out the procedures described below without research or evaluation staffs, without computer facilities, and with little or no technical assistance.

Example A

District A constructed a composite pretest measure consisting of pretest scores and teacher ratings which was used for selecting students and estimating project gains in a special regression model design. The Stanford Achievement Test was administered to grades 2 through 6 in the spring. Classroom teachers were asked to rate the reading ability of each student on a five point scale. The rating was scored 10, 30, 50, 70, or 90 to give it a range of scores similar to the pretest. Since the district wished to weight pretest scores more highly than ratings, the composite score was computed by the formula:

$$C = 1.4X + .6R$$

where C is the individual's composite score, X is the pretest percentile, and R is the teacher rating.

The means, standard deviations, and correlations for pretest, rating, composite and posttest are given in Table 1. The rating correlates highly with the pretest and both variables correlate highly with posttest scores. The composite score did not correlate significantly higher with posttest scores than did either of these variables separately.

It is interesting to note that the scaling of the teacher rating produced a score distribution that was close to normal for each grade. Means and standard deviations for the rating were very close to what one would expect for NCEs, 50 and 21.06, respectively. Thus, the rating could simply be added to the pretest.

As might be expected, the results of an analysis of covariance suggested that there were significant differences in the mean ratings by various teachers. Pretest scores were used as the covariate to partially correct for the achievement level of the class. The high correlations observed above, though, suggest that these inter-rator differences did not seriously affect the validity of the ratings.

Example B

District B administered the Total Reading and Total Mathematics subscales of the Stanford Achievement Test each spring to evaluate the reading and mathematics programs at grades 2 and 3. Although the district had been using the norm referenced evaluation model, there was some evidence that the equipercentile assumption did not hold for their students. In the primary grades, students consistently performed well compared to the national average, but the class average droped in each successive year until it approached the national average at about grade five.

Since students were selected using teacher judgment, the data were not appropriate for analysis using the special regression model, but the district did collect ratings and other information from which composites were computed. While these composites were available to the teacher when students were assigned to Title I, they were only used to set a very unrestrictive upper limit above which students could not be selected for Title I services. Thus, the match between composite scores and teacher judgments could be explored using actual referrals coded 1 to indicate Title I instruction and 0 to indicate no supplemental instruction.

The composite was constructed during the fall. Classroom teachers were asked to rate each student in seventeen reading, mathematics and study skill areas on a four point scale. In mathematics an informal inventory developed by the district was also administered. In reading, students were assigned points on the basis of the grade placement of his or her current level in the reading series. An attempt was made to equate the difficulty of levels of the three different reading series used across buildings. Each element of the composite; pretest, rating and informal inventory or reading placement; was then rescaled by assigning points to each score so that the rescaled values on each score ranged from 0 to 15. The points were simply totaled to form the composite score.

Unfortunately the point system was devised without looking at score distributions, and as a result, rather extreme floor and ceiling variables were observed on each of the variables. Figure 1 shows the score distributions for each variable and the composite totals for grade 3 reading. By way of comparison, the original pretest distribution in Normal Curve Equivalents (NCEs) is also provided. Although the raw scores for reading placement and the informal inventory were unavailable, their distributions would probably roughly approximate normality. The rating scale, on the other hand, had been scaled by dividing the raw scores by a constant and rounding to a whole number. An explanation for the ceiling effect observed on this variable was suggested by district officials. Teachers probably had a response set to overrate the needs of students due to the length of the rating scale and thus ensure that certain students would not be deprived of Title I services.

The means, standard deviations and correlations for each of the composite and criterion variables are given in Table 2 for students in reading. Correlations among the composite variables were moderate while correlations with the composite were high. Correlations of this variables with the criterion variables were moderate to high.

The composite correlated highly with the posttest, but not significantly higher than the pretest alone. The composite correlated very highly with referrals, significantly higher than the pretest alone. These results suggest that while the use of composites would not increase the precision of the special regression model, a closer match to the subjective selection procedure is possible.

Inter-rator differences were again observed across teachers, but these did not appear to seriously affect the results.

Table 3 presents the means, standard deviations and correlations among the composite and criterion variables for students in mathematics. The results were similar to those observed in reading except that at grade 3 the composite did not correlate more highly with teacher judgment than the pretest alone.

<u>Example C</u>

With the help of an outside contractor, District C evaluated its Title I reading program in grades 1 through 6 using the special regression model. Students were selected with a single score cutoff on a composite of teacher ratings and scaled scores on the Total Reading subscale of the Stanford Achievement Test. Each spring, classroom teachers rated students on eight reading and study skills with a five point scale and administered the achievement test. The composite was constructed using the recommended procedure for multiplying one of the variables by a constant to equate the standard deviations and summing the scores.

Although it was only possible to examine the scores from one elementary school, the data from District C are of special interest since two independent ratings of the same students by different teachers were available. One rating was made near the administration of the pretest and the other was made a year later. This provided the opportunity to study the consistency of rators and to examine the effects of summing the two ratings.

The means, standard deviations and correlations for the pretest, teacher ratings, summed ratings, and posttest are presented in Table 4. Again the correlation between the ratings and test scores are high. The composite correlates highly with the posttest but no higher than the pretest alone.

The correlation between the two ratings is remarkably high considering that different teachers made the ratings a year apart. When the ratings were combined to increase reliability, higher correlations with test scores were observed but the sample was too small to detect significant differences.

Example D

The fourth case will be described only briefly due to the small amount of data available and certain problems in the procedures used. Pretest scores, teacher ratings and selection for Title I were available for second grade students only. Ratings were obtained as part of a district feasibility study by asking the classroom teacher to estimate each students' reading ability on a six point scale. The directions for the scale encouraged the teacher to make use of test scores or informal inventories to help make their judgments. It is not known whether teachers actually made use of test scores, however.

The means, standard deviations and correlations for the pretest, rating and teacher referrals are presented in Table 5. The high correlation between the pretest and rating is not surprising if teachers did use test scores as the basis of their rating. However, the ratings correlated significantly higher with referrals than did the pretest. When a composite was formed using least squares regression coefficients, ratings added significantly to the prediction of teacher placement decisions. These findings suggest that the ratings reliably measured some aspect of teacher judgments not captured by test scores.

DEVELOPING COMPOSITES FOR STUDENT SELECTION

The four examples of field applications of composite scores described here have implications for those interested in developing composites to select Title I students. The variation in the way each district approached the task provides some basis for generalization, though clearly these data are not without limitations.

Selecting variables for the composite

The specific procedures used to select students for Title I services vary considerably from district to district and are usually unsystematic. When teacher judgment is the basis for Title I referrals, one or more of the following classes of variables generally enter into the decision:

<u>Performance indicators</u>--Indicators of achievement include test scores, grades, informal inventories, progress through a reading series, previous participants in Title I and classroom performance.

<u>Study skills</u>--Teachers generally look for attentiveness, work efficiency, ability to complete assignments and ability to follow directions.

<u>Non-academic factors</u>--Less directly related to achievement are such factors as motivation, personality and behavioral problems in class.

<u>Administrative problems</u>--There may be scheduling conflicts, overlapping services, parental concerns, or other administrative problems that affect the decision.

The task of choosing appropriate variables is not to be taken lightly. Clearly it would be impossible to include more than a small number of the variables listed above in a formal selection procedure.

with the exception of performance indicators, the variables would be difficult to measure. More important, the validity of many of these variables with respect to selecting the most "needy" students is questionable. Current performance in the subject matter area is, in fact, the criterion of primary interest to most districts, although other factors usually are considered as well.

Scores on standardized tests are generally a first choice for inclusion in a composite. They are reliable and valid indicators. Teachers, however, feel that test scores do not adequately measure a student's need for services and that their experience with students in the classroom enables them to make more valid placement of students. To quantify these teacher judgments, ratings are frequently used as in the four cases described here, yet ratings are notoriously subjective and unreliable (Thorndike & Hagan, 1977). How can ratings be expected to contribute to selection? Should other indicators be used?

<u>Teacher ratings</u>. Several different rating scales were described in the last section, from simple five-point scales to lengthy skills checklists. Despite the various procedures used, teacher ratings of student ability were shown to be surprisingly reliable and valid. Correlations with test scores, with teacher referrals (District B), and with independent ratings (District C) were consistently in the range of .60 to .75. Moderate correlations were observed with reading placement and an informal inventory in District B. These findings are consistent with other field studies (Hennessy, Takota & Ames, 1980).

With the high reliability of the pretest scores and the redundancy among all the achievement indicators, it is probably unreasonable to expect much improvement in predicting posttest scores by adding composites. On the other hand, teacher ratings do seem to reliably measure some factor that is used by teachers in selecting students but is not reflected in test scores. Thus by including ratings, the district should be better able to hold to a single score cutoff on the pretest composite. Teachers should be more satisfied with the referrals based on composites than pretest scores alone.

The inflated ratings and inter-rator differences observed by these districts suggest that care in collecting ratings is necessary. Clear instructions or training should be provided so that teachers understand how to make the rating, minimizing inflated ratings and ensuring a normal distribution of scores. Teachers should be told what percent of their students can be expected to fall under each rating, though this will vary with good and bad classes. It could be noted that inflated ratings reduce the variability or range of scores, thus reducing the contribution to the composite, and that students with less need may get selected over students with more need.

Probably one of the more effective ways to increase the reliability of a rating scale is to use more than one rater. Only a single elementary teacher will be familiar with a student's reading or math performance during the school year, but two ratings could be easily obtained if classroom teachers rated their students in the spring and again a couple weeks into the fall. While the teacher has had relatively little time to observe the student in the fall, reasonable ratings can be made as the District B results show. Treating the inter-rater correlations provided by District C as reliability estimates, we can predict the reliability expected from combining the two ratings using the Spearman Brown Prophecy Formula. If the reliability of ratings A and B are .70, the reliability of the combined ratings would be a sizable .82.

As the results in Table 4 suggest, the combined rating should correlate more highly with test scores.

In view of the high reliability of ratings in the examples described, inter-rator differences do not seem to be a major problem. The procedures described by Roberts and Tallmadge (1976) to correct for these differences are rather complicated to implement by the districts and are probably unnecessary. The procedure described above for obtaining two ratings on the same students would seem to be preferable.

Other performance indicators. Besides standardized test scores and teacher ratings, a number of other indicators of achievement were listed above that could be used in a composite. There would only be two reasons for including additional performance indicators: to increase the reliability of the composite and thus the correlation with posttest scores or to increase the validity of selection. The efforts of District B to include ratings, inventories and reading placement in composites had little effect on correlations with the posttest. On the other hand, this district had good reason to consider other indicators thought to more accurately reflect teacher judgments in student selection. Grade placement in the reading series correlated more highly with referrals than did teacher ratings.

Other factors. Not surprisingly, each of the districts focused only on performance measures. It could be argued that poor study skills are relevant to need but no data was available (with the exception of several items on District B's rating checklist). Administrative problems that are unrelated to level of achievement would not need to be included since students affected could simply be dropped from the evaluation. When forced to objectify the decision process, teachers omitted variables that might have been included otherwise but were difficult to defend.

Scaling the Variables

Once the variables have been chosen for the selection measure, numbers should be assigned to the values of each variable so that (a) the score distributions are normal, (b) the variables have the same standard deviation, and (c) the variables are scaled in the same direction. The problems encountered by the four districts and the solutions to them provide an interesting perspective on these issues.

Score distributions. Since the special regression model is sensitive to floor and ceiling effects, it is particularly important to inspect the score distributions in scaling the variables. The odd distributions displayed in Figure 1, for example, could have been avoided quite easily had the assignment of points reflected the full range of values on each variable. The close approximations to normality found for the other composites was rather encouraging, especially since one might have expected inflated scores on the ratings by teachers wishing to make sure students sufficient points to be selected.

Standardization. From the practitioner's perspective, standardizing the variables or multiplying by constants as recommended (Roberts and Tallmadge, 1976) is a nuisance. All but one district tried to approximate standardization by setting up equivalent scales on each variable. The rating scale used by District A, for example, yielded means and standard deviations remarkably close to that of the pretest NCEs. The point system constructed by District B achieved the same goal but with much tedious converting of scores.

The advantage of simple scales may be obvious when computer facilities are not available, but even when they are, the turnaround time required to generate the composites could delay the start of the program and manual computation may have to be used anyway.

Weighting the Composite

Assigning weights to each variable is the final step in constructing the composite. The weighting scheme used, though, is of secondary importance relative to variable selection and scaling. The evaluator can assign arbitrary weights to give one variable a larger contribution in the composite or find the least squares regression coefficients that maximize the correlation between the composite and some criterion using multiple regression. When the correlations between the variables and with the criterion are all of about the same magnitude, though, it is unlikely that unequal weights will substantially improve on unit or equal weights (Wainer, 1976).

To illustrate this point, unit weights and optimal least squares weights were compared for Districts A and B. The differential weights applied by District A (1.4, .6) were also examined. Multiple regression was used to find the least squares regression coefficients with posttest scores as the criterion. Since the least squares solution capitalizes on chance variation in the data, the Darlington (1968) formula was used to estimate the shrinkage that would be observed in the multiple correlations if the regression coefficients had been applied to a new set of data in a cross-validation design.

The squared correlations reported in Table 6 generally support the notion that equal weights work fine in most situations. There are only small differences between the weighting schemes. The optimal weights are slightly higher than district or unit weights but this difference disappears when the squared correlations are adjusted for estimated overfit to the data.

CONCLUSIONS

The special regression model requires that the treatment group be selected strictly on the pretest measure. School districts applying the model to Title I evaluation have found it difficult to meet this requirement when teacher judgments or other factors were added to the student selection process but not to pretest measure. This paper explored the feasibility of using composite scores to solve this problem.

As Dawes and Corrigan (1979) hare argued, composites make good models of decision making behavior. When the appropriate variables are included in the model, composite scores make excellent predictions of teacher referrals for Title I instruction. While the referrals process cannot be modeled exactly, a composite provides a better match than test scores alone and should make it easier for a district to enforce a strict cutoff. The district must examine its selection process to determine how it can best be modeled.

Contrary to expectations (Roberts & Tallmadge, 1976), composites do not seem to increase the accuracy of estimating the no treatment expectation in regression designs. At least with the data reported here, the pretest-posttest correlation was not increased by adding other performance indicators. Measures such as teacher ratings and placement in a reading series can be quite reliable but are redundant with pretest scores.

It is encouraging to find that quite simple procedures for constructing composites yield pretest measures that have all the properties necessary for the special regression model. A procedure requiring the simple sum of two or three variables has obvious practical advantages over elaborate score transformations or scoring schemes. At least in the western states, few districts have the resources for following the technical recommendations for designing composites to the letter.

Care in designing these simple procedures is required, though. If, for example, the scaling of a variable or score inflation on a rating scale causes floor or ceiling effects in the composite scores, problems are created when project gains are to be estimated with the special regression model.

Independent of the evaluation model used, evaluators and program specialists should have a concern for the validity of the process by which students are selected for compensatory education programs. Hennessy, Takata and Ames (1980) suggest that many "needy" students are not being served by Title I. While the definition of need used in the study requires some discussion, those authors are not alone in their assessment of current selection procedures. Composite scores provide an objective, observable alternative to the decision making processes commonly used.

As Dawes and Corrigan (1974) put it:

> ...an analysis of the tasks faced by the decision maker leads to the conclusion that linear models work well. It is, therefore, not surprising that linear models outperform intuitive judgment. Nor is it surprising that decision makers (insofar as they are behaving appropriately) are paramorphically well represented by linear models...The whole trick is to decide what variables to look at and then know how to add. (p. 105)

References

Darlington, R. B. Multiple regression in psychological research and practice, Psychological Bulletin, 1968, 69, 161-182.

Dawes, R. M. & Corrigan, B. Linear models in decision making, Psychological Bulletin, 1974, 81, 95-106.

Estes, G. D. & Anderson, J. I. Observed treatment effects with special regression evaluation models in groups with no treatment. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

Kenny, D. A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. Psychological Bulletin, 1975, 82, 345-362.

Hennessy, M., Takata, G., & Ames, N. L. Title I selection methods and the coverage of disadvantaged students. Evaluation Review, 1980, 4, 43-57.

Roberts, A. O. H & Tallmadge, G. K. Composite scores, ESEA Title I evaluation and reporting system (revised). Mountain View, CA: RMC Research Corporation, October 1976.

Tallmadge, G. K. Selecting students for Title I projects, ESEA Title I evaluation and reporting system. Mountain View, CA: RMC Research Corporation, January 1978.

Tallmadge, G. K. & Wood, C. T. User's guide, ESEA title I evaluation and reporting system (revised). Mountain View, CA: RMC Research Corporation, January 1978.

Thorndike, R. L. & Hagan, E. P. Measurement and evaluation in psychology and education. New York: John Wiley and Sons, 1977.

Wainer, H. Estimating coefficients in linear models: It don't make no nevermind. Psychological Bulletin, 1976, 83, 213-217.

Yap, K., Estes, G. & Hansen, J. Effects of data analysis methods and selection procedures in regression models. Paper presented at the AERA annual meeting, San Francisco, 1979.

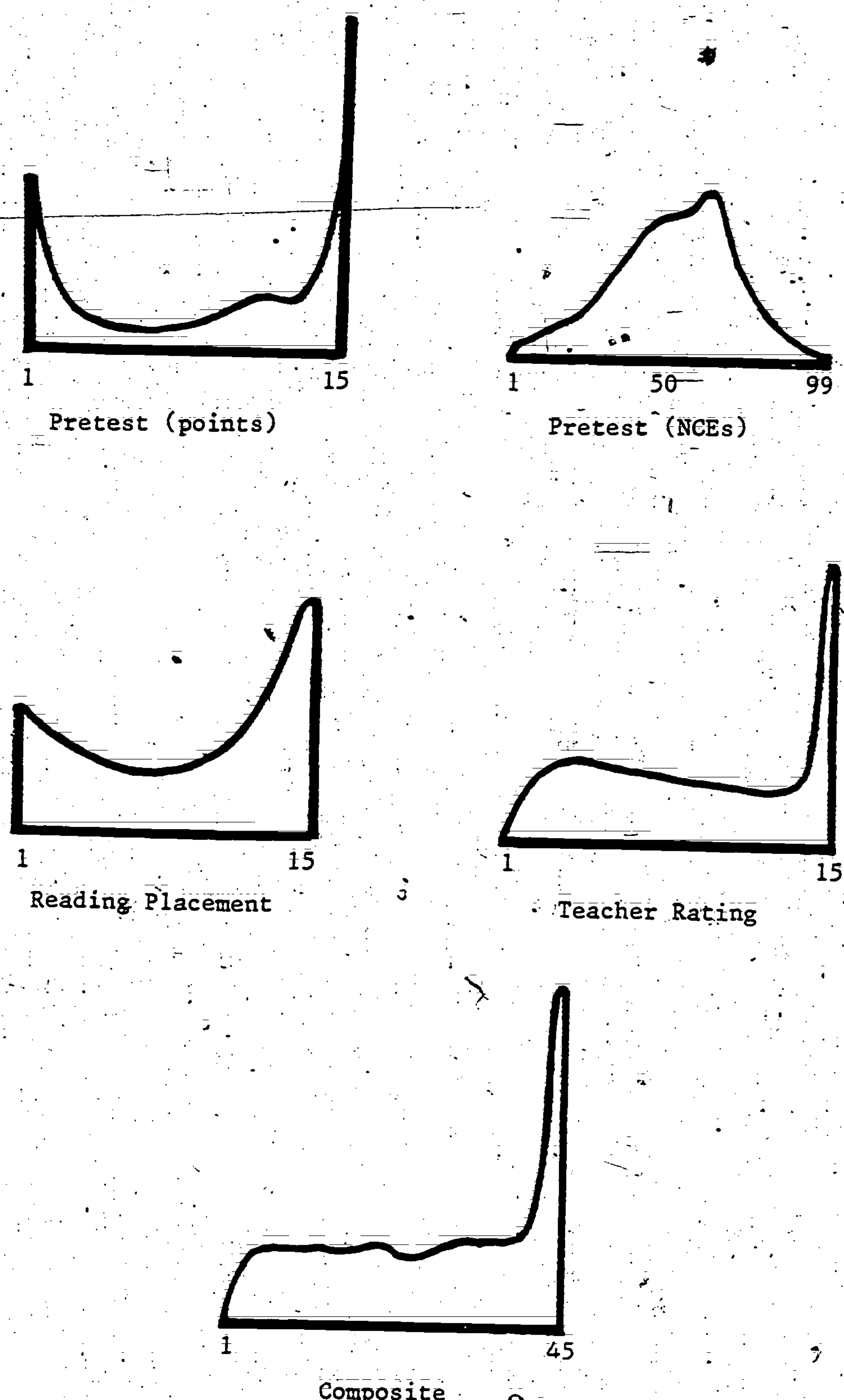

Figure 1. Score distributions on composite variables and composite total for third grade students in Reading from District B (N = 168).

Table 1
Means, Standard Deviation and Correlations
for District A Students in Reading

	X	SD	Correlations 1	2	3	4
Grade 2 (N = 897)						
(1) Pretest[1]	55.0	19.7				
(2) Rating	48.4	21.2	.75			
(3) Composite	109.1	50.2	.98	.84		
(4) Posttest	55.7	20.2	.75	.65	.76	
Grade 4 (N = 942)						
(1) Pretest[1]	55.2	17.8				
(2) Rating	49.6	20.7	.76			
(3) Composite	111.4	47.1	.98	.86		
(4) Posttest	51.4	19.2	.70	.66	.73	
Grade 6 (N = 855)						
(1) Pretest[1]	51.6	17.9				
(2) Rating	51.9	22.6	.74			
(3) Composite	104.3	47.6	.97	.86		
(4) Posttest	51.5	18.4	.75	.74	.78	

[1]Pretest and posttest means on scores are reported in Normal Curve Equivalents (NCEs).

Table 2
Means, Standard Deviations and Correlations
Among Composite Variables, Posttest and Treatment for
District B Students in Reading

	X	SD	Correlations 1	2	3	4	5
Grade 2 (N = 120)							
(1) Pretest[1]	8.3	6.3					
(2) Reading level	9.9	5.4	.74				
(3) Checklist	10.4	4.6	.55	.69			
(4) Composite	28.4	14.3	.89	.92	.82		
(5) Posttest[2]	59.9	20.9	-.69	-.65	-.58	-.73	
(6) Referrals	.62	.49	.70	.90	.61	.84	-.53
Grade 3 (N = 97)							
(1) Pretest	9.3	5.8					
(2) Reading Level	9.3	6.3	.68				
(3) Checklist	9.3	5.2	.58	.46			
(4) Composite	27.8	14.7	.88	.85	.79		
(5) Posttest[2]	54.9	18.4	-.79	-.63	-.49	-.76	
(6) Referrals	.65	.48	.77	.79	.64	.86	-.66

[1]Pretest scores were converted to a 15-point scale.

[2]Negative correlations reflect the reversed scaling of the composite variables.

Table 3
Means, Standard Deviations and Correlations Among Composite Variables, Posttest and Treatment for District B Students in Mathematics

	X	SD	Correlations 1	2	3	4	5
Grade 2 (N = 104)							
(1) Pretest[1]	8.2	6.1					
(2) Inventory	7.0	4.8	.60				
(3) Rating	10.5	4.4	.43	.46			
(4) Composite	25.9	12.7	.87	.83	.73		
(5) Posttest[2]	59.3	20.8	-.76	-.56	-.44	-.75	
(6) Referrals	.62	.49	.75	.75	.55	.84	-.58
Grade 3 (N = 96)							
(1) Pretest[1]	8.9	6.3					
(2) Inventory	12.2	3.7	.64				
(3) Rating	9.3	5.3	.54	.56			
(4) Composite	30.3	12.9	.88	.82	.83		
(5) Posttest[2]	58.2	20.9	-.72	-.56	-.49	-.70	
(6) Referrals	.69	.46	.85	.65	.60	.83	-.64

[1]Pretest scores were converted to a 16-point scale.

[2]Negative correlations reflect the reversed scaling of the composite variables.

Table 4
Means, Standard Deviations and Correlations for Composite Variables and Posttest for District C Students in Reading

	X	SD	Correlations 1	2	3	4	5
Grade 2 (N = 46)							
(1) Pretest[1]	121.9	15.0					
(2) Rating A	20.3	5.1	.75				
(3) Composite	242.9	34.3	.85	.84			
(4) Posttest[1]	137.6	12.8	.67	.78	.72		
(5) Rating B	24.5	5.3	.68	.78	.71	.73	
(6) A + B	44.8	9.8	.76	.94	.82	.80	.95
Grade 4 (N = 45							
(1) Pretest [1]	142.9	17.5					
(2) Rating A	23.2	6.7	.45				
(3) Composite	285.7	29.8	.88	.76			
(4) Posttest[1]	153.1	18.2	.75	.54	.74		
(5) Rating B	24.6	8.7	.77	.72	.85	.78	
(6) A + B	47.8	14.3	.68	.90	.87	.72	.95
Grade 6 (N = 44)							
(1) Pretest[1]	158.3	18.4					
(2) Rating A	22.1	6.6	.69				
(3) Composite	317.0	34.9	.92	.87			
(4) Posttest[1]	167.0	16.8	.82	.71	.81		
(5) Rating B	23.3	8.1	.75	.76	.80	.78	
(6) A + B	45.5	13.8	.77	.93	.89	.80	.95

[1] Pretest and posttest are reported in expanded scale scores.

Table 5
Means, Standard Deviations and Correlations Among Pretest Scores, Ratings and Teacher Referrals for District D Students in Reading (N = 242)

	X	SD	Correlations 1	2
(1) Pretest	57.8	21.0		
(2) Rating	3.7	1.4	.79	
(3) Refferals[1]	.28	.45	-.54	-.63

[1]The negative correlations reflect the reversed scaling of this variable.

Table 6
Squared Correlations Between the Posttest and Three Weighted Composites

Grade	N	District Weights	Unit Weights	Optimal Weights	Optimal Weights (Adjusted)
District A, Reading					
2	897	.581	.558	.582	.580
4	942	.534	.528	.535	.533
6	855	.626	.628	.631	.629
District B, Reading					
2	120		.533	.543	.515
3	97		.558	.629	.600
District B, Mathematics					
2	104		.571	.605	.577
3	96		.496	.541	.505