DOCUMENT RESUME

ED 087 437 IR 000 181

AUTHOR Nutt, A. T.; Batsell, R. R.
TITLE Multiple Linear Regression: A Realistic Reflector.
PUB DATE Apr 73
NOTE 21p.; Paper presented at the Association for Educational Data Systems Annual Convention (New Orleans, Louisiana, April 16 through 19, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Data Analysis; *Data Processing; *Decision Making; Decision Making Skills; *Educational Administration; Models; *Multiple Regression Analysis; Research Methodology
IDENTIFIERS AEDS; Association for Educational Data Systems; MLR; *Multiple Linear Regression

ABSTRACT

Examples of the use of Multiple Linear Regression (MLR) techniques are presented. This is done to show how MLR aids data processing and decision-making by providing the decision-maker with freedom in phrasing questions and by accurately reflecting the data on hand. A brief overview of the rationale underlying MLR is given, some basic definitions are offered, and MLR is described as the process of defining alternative linear models and using appropriate test statistics to determine the model which best satisfies the criteria of simplicity and goodness of fit. Following this, an actual question of interest to an educational decision-maker is stated, and a linear model which reflects the relevant factors is presented. The process of "testing the question" is outlined. An action-oriented interpretation of the data analysis and the resulting steps to be taken by the administrator are described. (Author/PB)

ED 087437

MULTIPLE LINEAR REGRESSION -

A REALISTIC REFLECTOR*

A. T. Nutt
R. R. Batsell

Texas Education Agency

PURPOSE

This paper should:

- enable the administrator to decide if he should investigate further, and become competent in, the use of linear models for decision-making.
- acquaint the administrator with analyses based on the use of linear models.

If the administrator already has a sufficiently strong background in statistics, this paper should equip him, with varying degrees of effectiveness, to perform the following tasks:

- express decision-questions in terms of expected-values.
- define a linear model for "testing the question of interest".
- arrive at an answer to the question of interest based on the appropriate test statistic.

In order to achieve these stated objectives, we simply want to give you some background definitions, familiarize you with linear models, and then share with you some of our experiences in using linear models for decision-making.

---

* The development of the technique reflected in this paper is largely the result of work accomplished by Bottenberg, Ward, and Jennings, and the approach is patterned after a graduate course offered by Earl Jennings at the University of Texas, Austin. Individuals enticed by this paper to pursue the topic further should obtain a copy of Introduction to Linear Models (Prentice Hall) by Ward and Jennings.

A. T. NUTT, as Director of the Division of Evaluation, Office of Planning, is responsible for evaluation of special projects and federally funded programs for the Texas Education Agency. He holds a B.B.A., a M.Ed., and is a candidate for the Ph.D in Educational Administration at the University of Texas in Austin.

RICHARD R. BATSELL is a programmer/analyst with the Management Information Center of the Texas Education Agency. He received a BBA in Statistics and a BA in Mathematics in 1971 and is currently studying Applications of Linear Models to decision-making at the University of Texas.

APPROACH

After the introductory rationale and definition of some necessary terms, this paper will adhere to the following structure:

- an actual question of interest to an educational decision-maker will be stated.
- a linear model that reflects the relevant factors will be presented.
- the process of "testing the question" will be outlined.
- action-oriented interpretation of the data analysis and the resulting steps taken by the administrator will then be described.

INTRODUCTORY RATIONALE

Data processing has established its role in managerial decision-making. For the education oriented community it has been the information vehicle for budget formulation, needs assessment, and program planning. Inherent in each of these areas is evaluation. Does spending money for X instead of Y make a difference in student achievement? Does group A need more help in reading than group B? Is curriculum C superior to curriculum D?

Administrators are making decisions daily which incorporate explicit or implicit answers to these questions. A relative measure of our success as processors of educational data is the extent to which the administrator does not have to compromise his inquiries to fit our processing tools. Ideally, he is a creative decision-maker asking meaningful questions about the intense human process of learning. Filled with interacting factors, this process strains the boundaries of our data systems (the information vehicle) more than an analogous process in any other environment. One tool capable of reflecting the human interactions involved in learning is "Multiple Linear Regression".

By "Multiple Linear Regression" we mean the process of defining alternative linear models and using appropriate test statistics to determine that model which best incorporates the criterion of "simplicity" and "goodness of fit".

PRELIMINARY DEFINITIONS

In order that the reader may be acquainted with some of the terms necessary to the development and use of linear models, we proceed with the following definitions.

expected value - the weighted sum of a set of values which yields an average for a group; it is this average that we call an expected value, i.e., "the expected reading level of third grade girls".

column vector - a columnar method of presenting a series of numbers;

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

is a column vector with three elements; if a column vector contains only ones and zeros it is called a binary vector.

linear model - an equation, designed to reflect relevant factors in a process, and composed of a series of vectors and their respective coefficients (also called weights); linear simply refers to the fact that the coefficients exist in our equation as multipliers and not as exponents.

criterion variable - the variable of interest that we hypothesize is dependent on the factors included in our model.

DESCRIBING THE PROBLEM

In order to proceed with our definitions we must now introduce an example of a linear model. Suppose we had scores from the administration of a reading test to a group of four students. Subsequent to the testing, the four students were randomly assigned to one of four teaching machines. Two of the machines were programmed with teaching method A and two with teaching method B. After an appropriate amount of instruction time they were then given a second administration of the reading test and their "gain score" was computed. We now want to determine if the two teaching methods are different in their effectiveness.

DEFINING THE STARTING MODEL

Entertain the following linear model:

$$Y = a_1A + a_2B + E$$

$$\begin{bmatrix} 1 \\ 3 \\ 2 \\ 6 \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

- the Y vector contains the measures on our criterion variable (the "gain score") for each of the four students.
- the A vector contains a representation for membership in the group of students taught by method A; specifically, it contains a 1 if the corresponding "gain score" in vector Y was observed on a person that was taught with method A; a 0 otherwise. (Note that A is a binary vector.)
- the B vector contains a representation for membership in the group of students taught by method B; specifically, it contains a 1 if the corresponding "gain score" in vector Y was observed on a person that was taught with method B; a 0 otherwise. (Note that B is a binary vector.)
- the E vector (called the error vector) contains the numbers necessary to satisfy the equality once the values of a_1 and a_2 (the vector weights) are chosen; these numbers may be thought of as the deviations of each student's actual score from the expected value of a group in which he is a member.

We know that the first two elements of Y are measures on students taught by method A and the last two, measures on students taught by method B. From our model, the elements in Y can be expressed by the equations:

$$
\begin{aligned}
1 &= a_1\,(1) + a_2\,(0) + b_1 \\
3 &= a_1\,(1) + a_2\,(0) + b_2 \\
2 &= a_1\,(0) + a_2\,(1) + b_3 \\
6 &= a_1\,(0) + a_2\,(1) + b_4
\end{aligned}
$$

SOLVING FOR THE VECTOR WEIGHTS

It should be obvious that there are a number of different values for a_1 and a_2 (our vector weights) that would satisfy the equations. But which values shoud we use? In order to answer the question let's begin by experimenting with different values:

$$
\begin{aligned}
1 &= 1\,(1) + 1\,(0) + (0) \\
3 &= 1\,(1) + 1\,(0) + (2) \\
2 &= 1\,(0) + 1\,(1) + (1) \\
6 &= 1\,(0) + 1\,(1) + (5)
\end{aligned}
\qquad \text{so} \qquad E_1 = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 1 \\ 5 \end{bmatrix}
$$

let $a_1 = 2$ and $a_2 = 4$

$$
\begin{aligned}
1 &= 2\,(1) + 4\,(0) + (-1) \\
3 &= 2\,(1) + 4\,(0) + (1) \\
2 &= 2\,(0) + 4\,(1) + (-2) \\
6 &= 2\,(0) + 4\,(1) + (2)
\end{aligned}
\qquad \text{so} \qquad E_2 = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ -2 \\ 2 \end{bmatrix}
$$

let $a_1 = 4$ and $a_2 = 1$

$$
\begin{aligned}
1 &= 4\,(1) + 1\,(0) + (-3) \\
3 &= 4\,(1) + 1\,(0) + (-1) \\
2 &= 4\,(0) + 1\,(1) + (1) \\
6 &= 4\,(0) + 1\,(1) + (5)
\end{aligned}
\qquad \text{so} \qquad E_3 = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} -3 \\ -1 \\ 1 \\ 5 \end{bmatrix}
$$

Note that the values for b_1, b_2, b_3, and b_4 (the elements of the error vector) depend on the chosen values of a_1 and a_2.

In order to specify the basis for determining the "best possible values" for a_1 and a_2 (the vector weights) we introduce the following definitions:

least squares solution - values for the vector weights that cause the sum of the squares of the elements in the error vector (E) to be a minimum; in our specific example, values for a_1 and a_2 such that no other values would yield a smaller result of $b_1^2 + b_2^2 + b_3^2 + b_4^2$.

error sum of squares - the sum of the squares of the elements in the error vector.

When the error sum of squares for a model is a minimum, the values for the vector weights are said to be the result of a least squares solution.

From our choices for a_1 and a_2 we find

- when $a_1 = 1$ and $a_2 = 1$, $(0)^2 + (2)^2 + (1)^2 + (5)^2 = 30$
- when $a_1 = 2$ and $a_2 = 4$, $(-1)^2 + (1)^2 + (-2)^2 + (2)^2 = 10$
- when $a_1 = 4$ and $a_2 = 1$, $(-3)^2 + (-1)^2 + (1)^2 + (5)^2 = 36$

Of these three sets, $a_1 = 2$ and $a_2 = 4$ cause the sum of the squares of the elements in E (the error sum of squares) to be the smallest. In fact, there are no other values for a_1 and a_2 that produce a smaller error sum of squares. (There are a number of computer programs that will take vectors representing a linear model and produce a least squares solution for the respective vector weights.)

In this specific case (with the least squares solution for a_1 and a_2) one might note that the average "gain" (expected value) for the sample group taught by method A is equal to the value for a_1 and the average "gain" (expected value) for the sample group taught by method B is equal to the value for a_2. (This fact is characteristic of models with multiple groups defined by binary predictor vectors.)

From this point on, any reference to vector weights will mean weights which, for that model, produce the minimum error sum of squares.

Now back to our question at hand - are the teaching methods different?

In our model the expected value for method A is a_1 and each student's actual score differs from that expected value by the corresponding amount in the error vector. The same is true for method B and a_2. This verifies the fact that our first model allows for two different expected values for the "gain scores" - a_1 for method A and a_2 for method B. And finally, in our first model with the least squares solution for a_1 and a_2 as 2 and 4 respectively, our error sum of squares is 10.

DEFINING THE RESTRICTED MODEL

Now let's construct a model that does not allow for differences between the two expected values, find the least squares solution for any vector weights that model contains, and then look at the error sum of squares. To generate this second model, we must impose the restriction that $a_1 = a_2$; we are in effect requiring that the expected values for the two groups be equal. With this restriction our model becomes

$$Y = a_1 A + a_1 B + E$$

(Note that since we have required that $a_1 = a_2$ we can substitute a_1 for a_2 as the coefficient for the B vector.)

Since $a_1 A + a_1 B = a_1(A + B)$ we have

$$Y = a_1(A + B) + E$$

$$\begin{bmatrix} 1 \\ 3 \\ 2 \\ 6 \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} \quad \text{or} \quad \begin{aligned} 1 &= a_1 (1) + d_1 \\ 3 &= a_1 (1) + d_2 \\ 2 &= a_1 (1) + d_3 \\ 6 &= a_1 (1) + d_4 \end{aligned}$$

(Note that there is only one expected value (a_1) in this new model.)

The least squares solution for the vector coefficient (a_1) in this new model is 3, and $d_1 = -2$, $d_2 = 0$, $d_3 = -1$, $d_4 = 3$ or

$$E_4 = \begin{bmatrix} -2 \\ 0 \\ -1 \\ 3 \end{bmatrix},$$

and finally, the error sum of squares is computed $(-2)^2 + (0)^2 + (-1)^2 + (3)^2 = 14$.

At this point it would be helpful to summarize at an intuitive level:

We want to find out if the two teaching methods produce different results. We first define a linear model that allows different expected values for the variable of interest. We then find the "best possible estimates" for those expected values based on the data at hand. (The "best possible estimates" are defined as those that cause the error sum of squares to be a minimum.)

In the next step we define a model that does not allow for differences in the expected values. The least squares solution for the vector weight in this "restricted model" is the "best possible estimate" for a common expected value. (No differences between expected values for the methods have been allowed.)

If the two expected values in the first model were almost equal, then the "restriction" to a common value will yield a model that "fits" almost as well. Therefore, the difference in the measure of "fit" (the difference in the error sum of squares) from the first model to the second will not be "very large".

On the other hand, if the two expected values in the first model were "very different", the "restriction" to a common value will yield a model that has an error sum of squares that is "much larger".

TESTING FOR SIGNIFICANCE

The basis for deciding what constitutes "very different" or "much larger" follows. Assume the data is a representative sample randomly drawn from the population of interest. It is possible for us to have drawn our sample in such a way that the students assigned to method A actually averaged less than would have all students taught by A; similarly, the students taught by method B may have performed better than the average we would have encountered had all students used method B. Therefore, we must "make relatively certain" that the sample performance differences between the methods are due to real differences in the effect of the methods and not simply a result of our randomizing process. To "make relatively certain" is called "testing for significance".

In order for us to "test for significance" the following assumptions about our data should be verified:

1) Our sample should be randomly selected from the population of interest.
2) It must be possible to express the actual expected values of the population in the manner of our first model.
3) The distribution of the variable of interest within each population (group) must be normal.
4) The variance of the variable of interest must be the same within each population (group).
5) The variable of interest within each population is distributed independently of any of the other populations.

(Although full understanding and careful application of these assumptions is crucial to significance testing, we do not think such understanding is necessary to the flow of this paper.)

If it is true that the data possesses the preceding characteristics, then we may proceed with an F-test. An F-test involves the calculation of the F-statistic which is determined by the following formula:

$$F = \frac{(ESS_R - ESS_F)/df_1}{ESS_F/df_2}$$

where

- ESS_R is the error sum of squares for the second or "restricted" model.
- ESS_F is the error sum of squares for the starting or "full" model.

- df_1 and df_2 are "degrees of freedom" (df_1 equals the number of linearly independent predictor vectors in the full model minus the number of linearly independent predictor vectors in the restricted model; df_2 equals the number of elements in the criterion vector minus the number of linearly independent predictor vectors in the full model*).

Temporarily ignore the effect of df_1 and df_2. Note that the numerator of the fraction is the difference between the error sum of squares for the restricted model and the error sum of squares for the starting (full) model; the fraction itself is simply the ratio of that difference to the error sum of squares for the starting (full) model. For any given full model, a comparison to restricted models that have larger and larger error sums of squares will result in larger and larger values for F. When the value for F exceeds some "prespecified" amount, the differences in the effectiveness of the representation of the data by the models is said to be significant. But since we know that the first model incorporates all necessary expected values, it follows that the restrictions resulting in the second model do not adequately allow for the reflections of relationships that exist in the data.

Each significance test results in an F-value. There is always a possibility that the calculated F-value reflects differences found in the sample groups, while no differences actually exist in the two populations. Standard tables for the F-distribution contain the probability of such an error. It is this probability value which the administrator must ultimately incorporate into his decision-making.

At this point we should mention that significance testing is applicable when the data is a random sample. If, as in some situations, data exists on the entire population of interest then we recommend a simpler procedure. Construct a starting model which reflects possible differences in expected values; find the least squares solution for the expected values; and finally, calculate the differences in the expected values. The administrator can then base his decision on the size of the difference. In our specific example, if the decision-maker feels that an average "gain score" of 4 instead of 2 is worth whatever other differences between method A and method B exist then he has reached his conclusion. If he does not feel the difference in the expected values warrants the recommendation of one method over another then he has also reached his conclusion.

ALTERNATIVE STARTING MODELS

We have now established all of the terms and definitions necessary for understanding the remainder of the paper. In the next step of the background and definition phase we discuss a different starting model for the example problem of the teaching methods. The following discussion should highlight the flexibility of linear models in reflecting relevant factors.

Suppose we show a colleague the starting model

$$Y = a_1A + a_2B + E$$

where the vectors are defined as before. Upon realizing that we used "gain scores" further suppose that he asked to see the original scores upon which the calculation of the "gain scores" was based. They are presented below:

* Full understanding is not necessary to the flow of the paper.

Student	Pre-test	Post-test	Gain
1	8	9	1
2	6	9	3
3	6	8	2
4	2	8	6

After examining the table he points out that teaching method B "had a lot more to work with" because students assigned to it scored so low on the pre-test, and that "it's not fair to compare the two methods using the 'gain score'". Further, suppose he adds that he's "not interested in a 'gain score' anyway". He wants to know "where the student is" when the instruction period is over. We acknowledge that his concerns are valid and begin thinking about alternative models which will incorporate the essence of his objections and still allow the "testing of the question of interest".

One possible such model could contain the post-test as the criterion variable and include the pre-test as one of the predictor variables. It is shown below:

$$Y = a_1A + a_2B + a_3PA + a_4PB + E$$

$$\begin{bmatrix} 9 \\ 9 \\ 8 \\ 8 \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} + a_3 \begin{bmatrix} 8 \\ 6 \\ 0 \\ 0 \end{bmatrix} + a_4 \begin{bmatrix} 0 \\ 0 \\ 6 \\ 2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}.$$

- the Y vector contains the post-test score for each student.
- the A vector contains a 1 if the post-test score in Y was observed on a student taught by method A; a 0 otherwise.
- the B vector contains a 1 if the post-test score in Y was observed on a student taught by method B; a 0 otherwise.
- the PA vector contains the pre-test score (for students taught by method A) that corresponds with the post-test score in Y; a 0 for students taught by method B.
- the PB vector contains the pre-test score (for students taught by method B) that corresponds with the post-test score in Y; a 0 for students taught by method A.
- the E vector is the error vector.

Note that in this model the criterion variable is no longer the "gain score". It is now the post-test score. Furthermore, the expected values for the two groups now "depend on" membership in the group, and the beginning level of achievement as measured by the pre-test.

The expected value for a student taught by method A and who had a pre-test score of x can be expressed as:

$$E(A,x) = a_1(1) + a_2(0) + a_3(x) + a_4(0)$$

where a_1, a_2, a_3, and a_4 are least square vector weights. Coefficients multiplied by 0 drop out and the equation becomes

$$E(A,x) = a_1(1) + a_3x$$

For any given score on the pre-test we can now estimate how a student taught by method A will perform on the post-test. Following the same procedure we could estimate expected performance on the post-test for students taught by method B. Furthermore, restrictions "testing" different questions of interest could be imposed as in our previous example. However, we have explored the use of this alternative starting model simply to highlight the flexibility of linear models.

In summary for this section we make the following points:

- the decision-maker should be permitted to decide on the form of his criterion variable.
- the decision-maker should be permitted to include all factors he deems to be effecting change in that criterion variable.
- finally, he should be able to test any hypothesis about which he has sufficient data.

It has been our experience that Multiple Linear Regression goes a long way in meeting these needs. In our next section we will share with you some of these experiences.

APPLICATION OF MULTIPLE LINEAR REGRESSION

PROBLEM SETTING

For many years kindergarten experiences have been available to those students whose parents could afford to enroll them in private schools. The benefit of such programs were apparently of value to those who bore the cost. It was said by many that this early childhood experience enabled these youngsters to accelerate their achievement in later school experiences.

With the coming of President Johnson's War on Poverty, this concept of the value of kindergarten was proffered as a way to enhance the educational attainment of those who were disadvantaged. Almost immediately the Office of Economic Opportunity initiated the Head Start programs. This was followed in the summer of 1966 with early childhood programs funded with Title I, ESEA. These programs were advocated by early childhood specialists and were sold by all of the agencies managing the poverty programs as a way of compensating for the lack of development experiences on the part of children about to enter the formal educational system.

In 1969 the Sixty-first Legislature of Texas enacted a sweeping educational law which provided, among other things, for the gradual phasing in (by 1975) of kindergarten type experiences for all youngsters age five and over. Few studies have been conducted to determine the actual long-term benefit of early childhood educational programs on the achievement of either disadvantaged or advantaged students. An evaluation of an early childhood program in Fort Worth Independent School District indicated that mean I.Q. scores of participants increased by as much as ten points during the one year (nine months). (3, pp. 91-93) An evaluation of such a program in Edgewood Independent School District indicates, "A significant gain at ages five, when tested for English language development." (4, p. 46)

A research study by Norman Silberberg to determine The Effect of Kindergarten Instruction in Alphabet and Numbers on First Grade Reading concluded that, "the beneficial effects of kindergarten training were dissipated by the end of grade one." (2, p. 14) Another study by Melvin Allerhand, which looked at the Head Start program, indicates that there are decreasing differences between the Head Start group and the non-Head Start group. (1, pp. 3-8)

Even without any clear evidence either to support or disclaim the value of early childhood education, the clamor to move early childhood education down to children ages two and three is widespread. There are, however, many who feel that this educational experience is not of much consequence and is an expensive baby-sitting service at best. As the convening of the Texas Sixty-third Legislature approached, the concern about early childhood education continued to be expressed. There was no clear cut mandate by the State Board of Education or the general public to do an evaluation study of early childhood education. However, we realized that some data was available to us which related early childhood education to some types of achievement data. It was our thinking that perhaps this available data would present at least some trends which would tend to support the experts in early childhood education.

The data which we had was collected by the 1971 Elementary School Survey from 98 Texas school districts. This survey was developed by the Joint Federal-State Task Force on Evaluation operating with funding from the U. S. Office of Education. The campuses which were surveyed were selected because of their participation in federally funded programs. A sample of pupils in Grades Two, Four, and Six on these campuses was included in the survey; however, their teachers provided the data needed to complete the data collecting instruments.

From these data, information was extracted about pupils who had experienced kindergarten or Head Start training and who had both pre- and post-test scores on some form of standardized test. Since the pupils included in this study were in either Grades Two, Four, or Six during the 1970-71 school year, their early childhood education was either through private programs or perhaps in some cases through programs funded by OEO or Title I, ESEA.

The implication of what I have said about the data up to this point, is that the sample was very biased by initial selection of the campuses and by subsequent scoring for those who had both pre- and post-test scores reported. The data is further contaminated by the gathering of all achievement scores into one pool regardless of the manufacturer of the test or the form of the test which was used. The fact that the early childhood programs were different for different groups of students may or may not be important to the findings.

THE PROBLEM

Does this data give any trend information which would support the theory that early childhood education is a factor in the performance of either disadvantaged or other students as reflected in achievement scores for reading or mathematics? If there is such a trend, does it have a greater impact on the performances of students who are disadvantaged?

There are a number of specific questions implied in the problem statement. Some of these are:

- Do second grade students who attended kindergarten perform better, as measured by the reading test, than second grade students who did not attend kindergarten?
- Do fourth grade students who attended kindergarten perform better, as measured by the reading test, than fourth grade students who did not attend kindergarten?
- Same for sixth grade.
- Do disadvantaged second grade students who attended kindergarten perform better, as measured by the reading test, than disadvantaged second grade students who did not attend kindergarten?

We found that each of these specific questions were "testable" using the same general model. Therefore, we will demonstrate the procedure for only three such questions. First consider the second graders.

DEFINING THE STARTING MODEL

Using post-test scores as the criterion and including factors reflecting

- the pre-test score,
- instruction time between tests, and
- kindergarten attendance,

we constructed the following starting model:

$$Y = a_1K + a_2NK + b_1IK + b_2INK + c_1TK + c_2TNK + d_1(IK*TK) + d_2(INK*TNK) + E$$

$$\begin{bmatrix} 2.4 \\ 1.8 \\ 3.6 \\ \vdots \\ 2.8 \\ 2.9 \\ 1.6 \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ 1 \\ 1 \end{bmatrix} + b_1 \begin{bmatrix} 1.8 \\ 1.4 \\ 2.1 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix} + b_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 2.1 \\ 2.4 \\ 1.7 \end{bmatrix} + c_1 \begin{bmatrix} 6 \\ 8 \\ 10 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 7 \\ 10 \\ 5 \end{bmatrix} + d_1 \begin{bmatrix} 10.8 \\ 11.2 \\ 21 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix} + d_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 14.7 \\ 24 \\ 8.5 \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

- the Y vector contains the post-test score for each student.
- the K vector contains a 1 if the corresponding score in the Y vector was observed on a student who had attended kindgergarten; a 0 otherwise.
- the NK vector contains a 1 if the corresponding score in the Y vector was observed on a student who had not attended kindergarten; a 0 otherwise.
- the IK vector contains, for all students who had attended kindergarten, the pre-test score corresponding to that post-test score in the Y vector; a 0 for all students who had not attended kindergarten.
- the INK vector contains, for all students who had not attended kindergarten, the pre-test score corresponding to that post-test score in the Y vector; a 0 for all students who had attended kindergarten.
- the TK vector contains, for all students who had attended kindergarten, the amount of instruction time between tests; a 0 for all students who had not attended kindergarten.
- the TNK vector contains, for all students who had not attended kindergarten, the amount of instruction time between tests; a 0 for all students who had attended kindergarten.
- each element in the IKTK vector is simply the product of the corresponding elements in the IK and TK vectors.
- each element in the INKTNK vector is simply the product of the corresponding elements in the INK and TNK vectors.

Note that the expected score on the post-test for a student that had attended kindergarten, scored 2.0 on the pre-test, and had eight months of instruction is expressed as:

$$E(K,2.0,8) = a_1 + b_1(2.0) + c_1(8) + d_1(16);$$

the expression for a non-kindergarten student with the same pre-test and instruction time is

$$E(NK,2.0,8) = a_2 + b_2(2.0) + c_2(8) + d_2(16)$$

After solving for the least squares vector weights, the error sum of squares was computed to be 219.94.

In order to test for significant differences between the performance of the students who had been to kindergarten and those who had not, we next imposed restrictions on the vector weights in the starting model that would force the expected values for the two groups to be the same. That is, let

$$a_1 = a_2,\ b_1 = b_2,\ c_1 = c_2,\ d_1 = d_2$$

The following model results:

$$Y = a_1(K + NK) + b_1(IK + INK) + c_1(TK + TNK) + d_1((IK*TK) + (INK*TNK)) + E$$

$$\begin{bmatrix} 2.4 \\ 1.8 \\ 3.6 \\ . \\ . \\ . \\ . \\ 2.8 \\ 2.9 \\ .6 \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ . \\ . \\ . \\ . \\ 1 \\ 1 \\ 1 \end{bmatrix} + b_1 \begin{bmatrix} 1.8 \\ 1.4 \\ 2.1 \\ . \\ . \\ . \\ . \\ 2.1 \\ 2.4 \\ 1.7 \end{bmatrix} + c_1 \begin{bmatrix} 6 \\ 8 \\ 10 \\ . \\ . \\ . \\ . \\ 7 \\ 10 \\ 5 \end{bmatrix} + d_1 \begin{bmatrix} 10.8 \\ 11.2 \\ 21 \\ . \\ . \\ . \\ . \\ 14.7 \\ 24 \\ 8.5 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ . \\ . \\ . \\ y_n \end{bmatrix}$$

Note that in our restricted model we have included the same relevant factors contained in the starting model, but we have not allowed for differences between students who attended kindergarten and those that did not. From this restricted model the expected value for any student with a pre-test score of 2.0 and eight months of instruction is expressed as:

$$E(2.0,8) = a_1 + b_1(2.0) + c_1(8) + d_1(16).$$

After solving for the least squares vector weights, the error sum of squares was computed to be 226.67.

Calculation of the F-statistic yields an F-value of 3.268. The probability associated with this F is .0045. Thus, if the total population of second grade students who had not attended kindergarten can actually perform identical to those who did, only .4 percent of the time would differences between sample groups be as large as those in our sample.

Using identical starting and restricted models, and samples of fourth and sixth graders, we produced table 1 shown below:

GRADE	ERROR SUM OF SQUARES, FULL	ERROR SUM OF SQUARES, RESTRICTED	F-STATISTIC	PROBABILITY
4	1601.56	1612.80	2.1053	.0780
6	761.75	787.90	4.8996	.0007

Table 1

The figures on the following pages were produced using the least squares weights generated by the appropriate starting model. For each of the two groups, at specific levels of performance on the pre-test, and different amounts of instruction time between tests, expected values for the post-test performance are plotted.

DECISION OPTIONS

This presentation, we hope, will illustrate for you the importance of multiple regression analysis in developing decision options which can consider various alternatives that may be controlled by the planner or evaluator. One could examine identified non-controllable variables, if these seem important; however, the planner or evaluator is usually interested in improving performance by describing more successful ways to mix the controllable variables.

In the example which we have presented, one of the controllable variables is early childhood education. This educational experience can be withdrawn, or increased amounts can be provided. In table 2 we examine this variable to see its impact as children move through levels two, four, and six. These are average gain scores for groups of students who have been exposed to early childhood education and for groups of students who have not been so exposed.

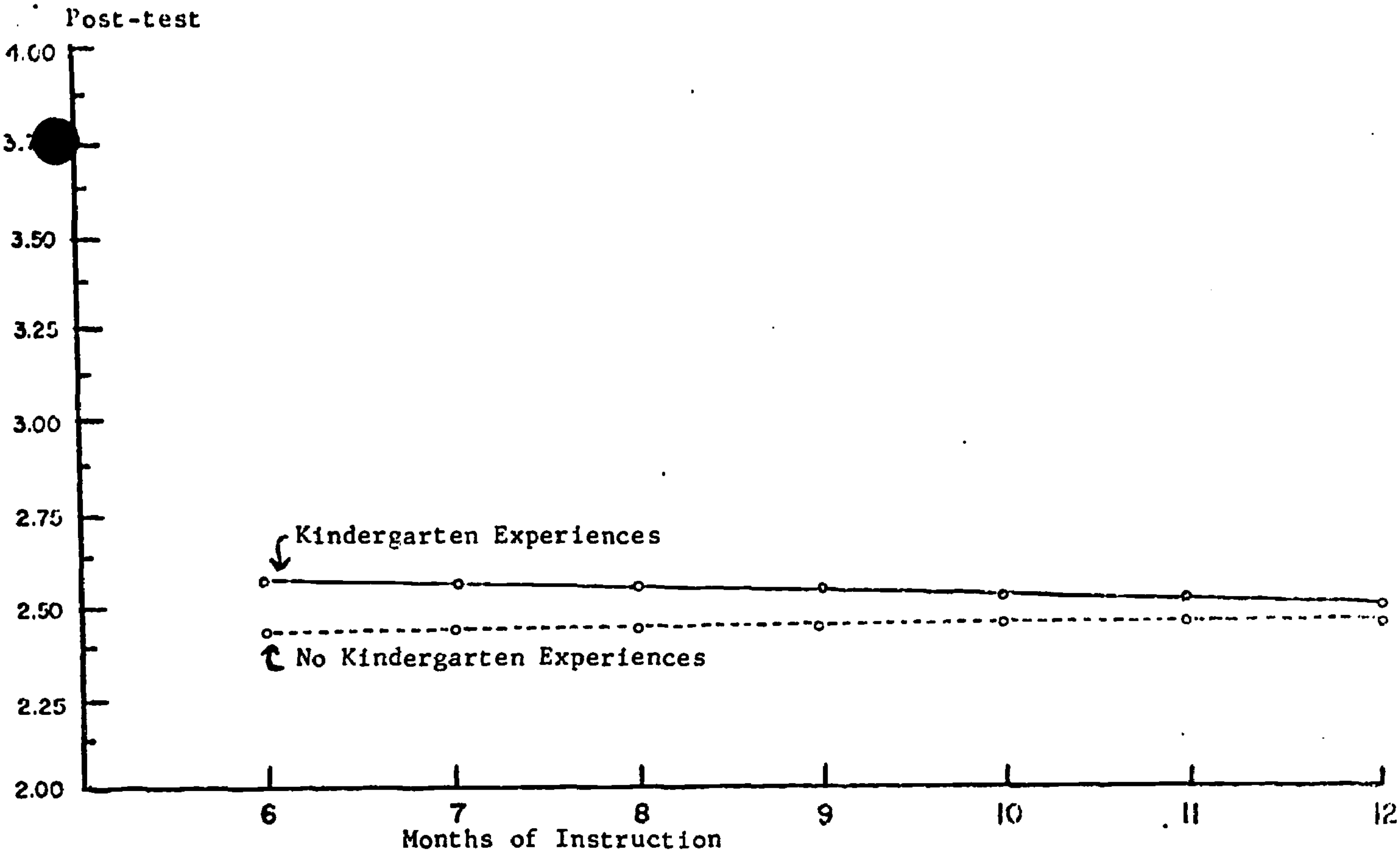

Figure 1 - Estimates of the Expected Values for Post-test Performance in reading by students scoring 1.6 on the Pre-test.

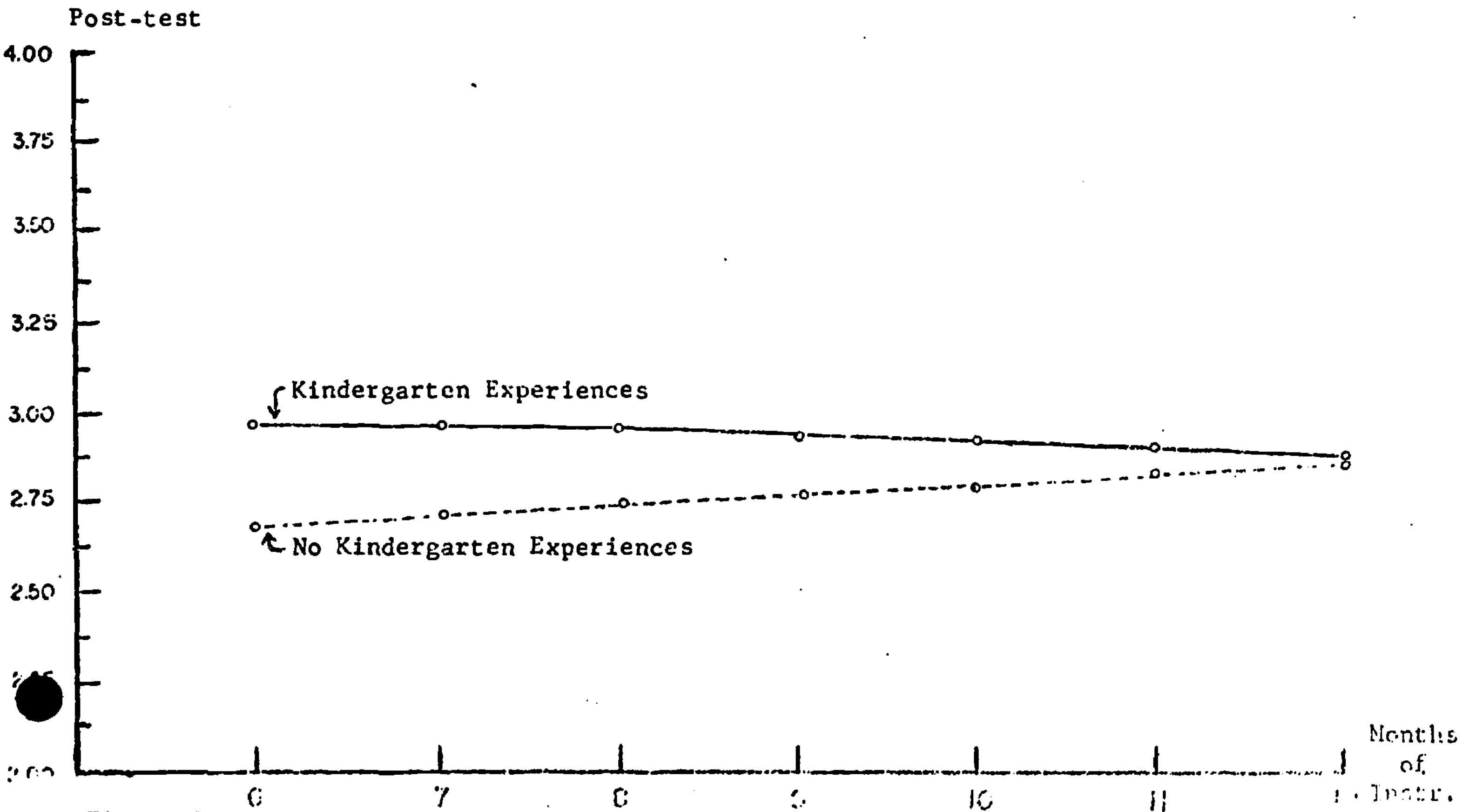

Figure 2 - Estimates of the Expected Values for Post-test Performance in reading by students scoring 2.0 on the Pre-test.

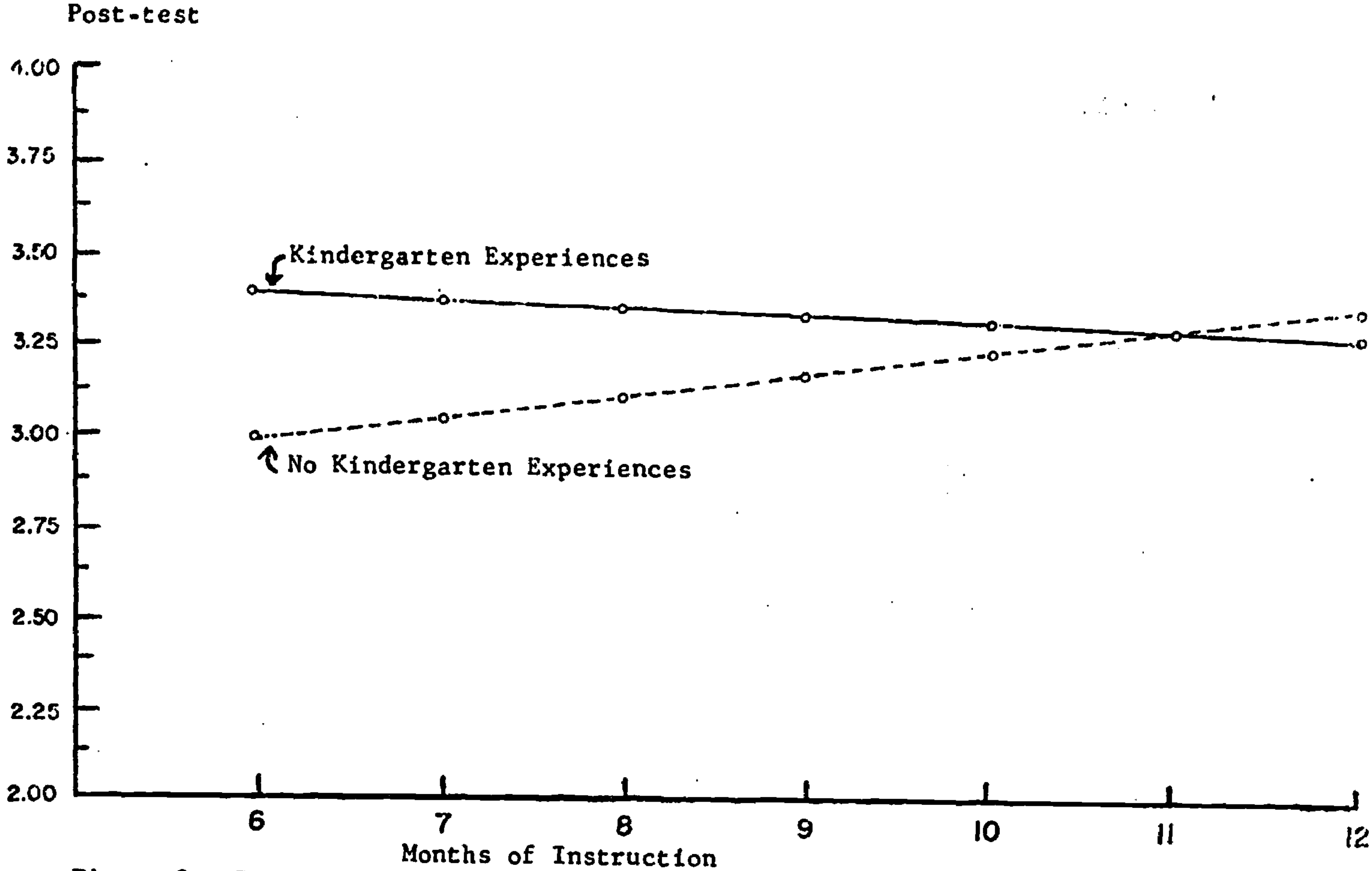

Figure 3 - Estimates of the Expected Values for Post-test Performance in reading by students scoring 2.4 on the Pre-test.

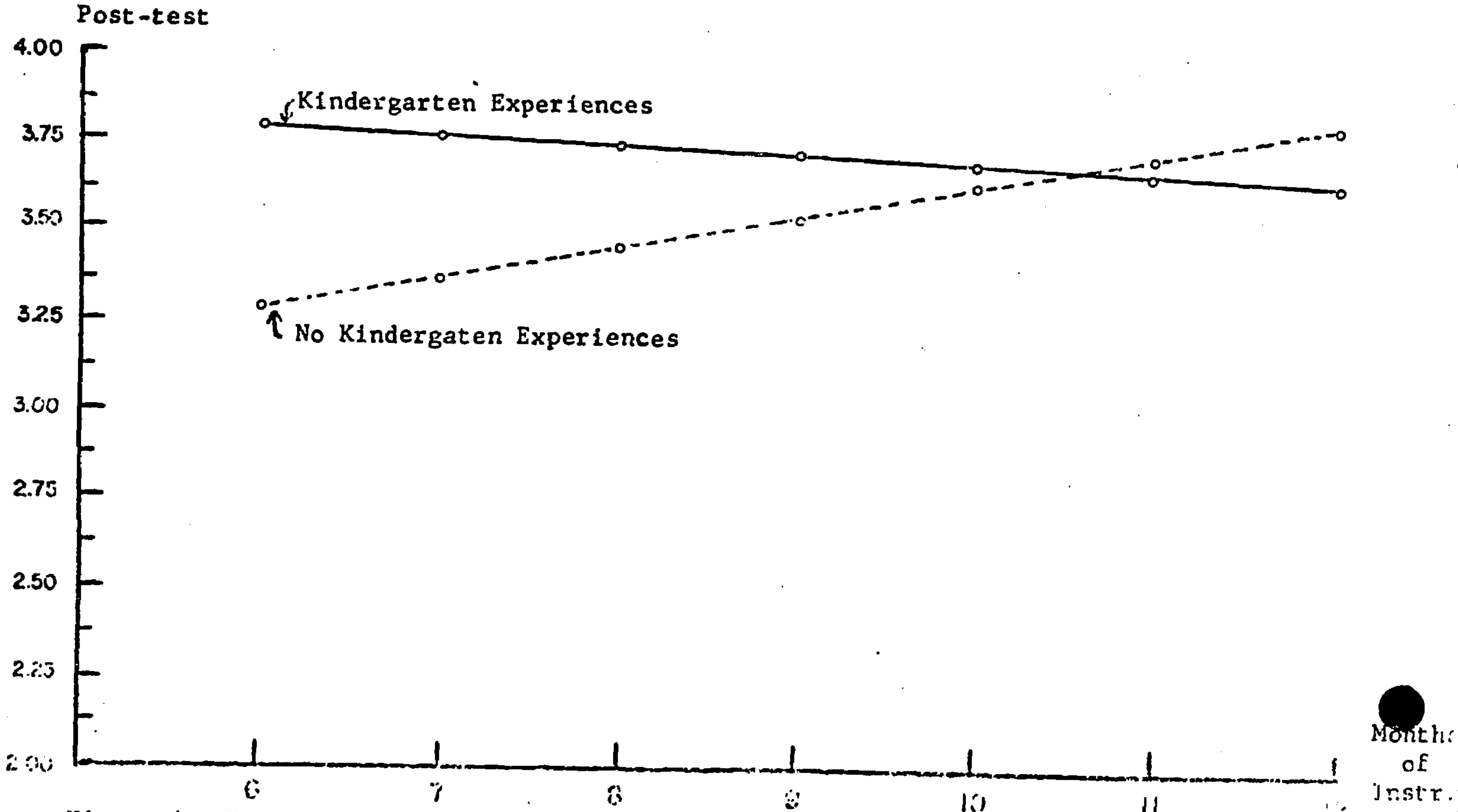

Figure 4 - Estimates of the Expected Values for Post-test Performance in reading by students scoring 2.8 on the Pre-test.

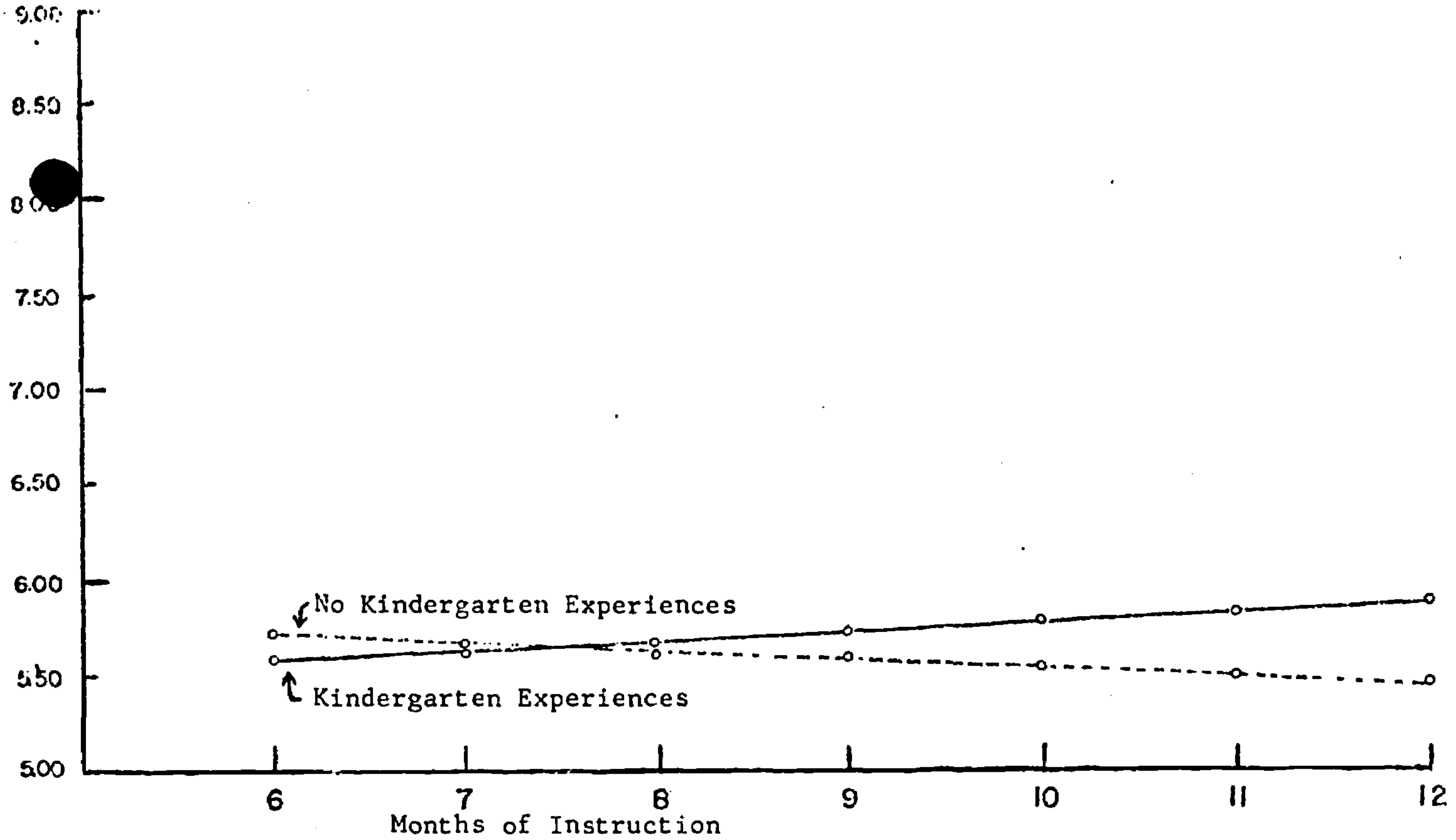

Figure 5 - Estimates of the Expected Values for Post-test Performance in reading by students scoring 4.8 on the Pre-test.

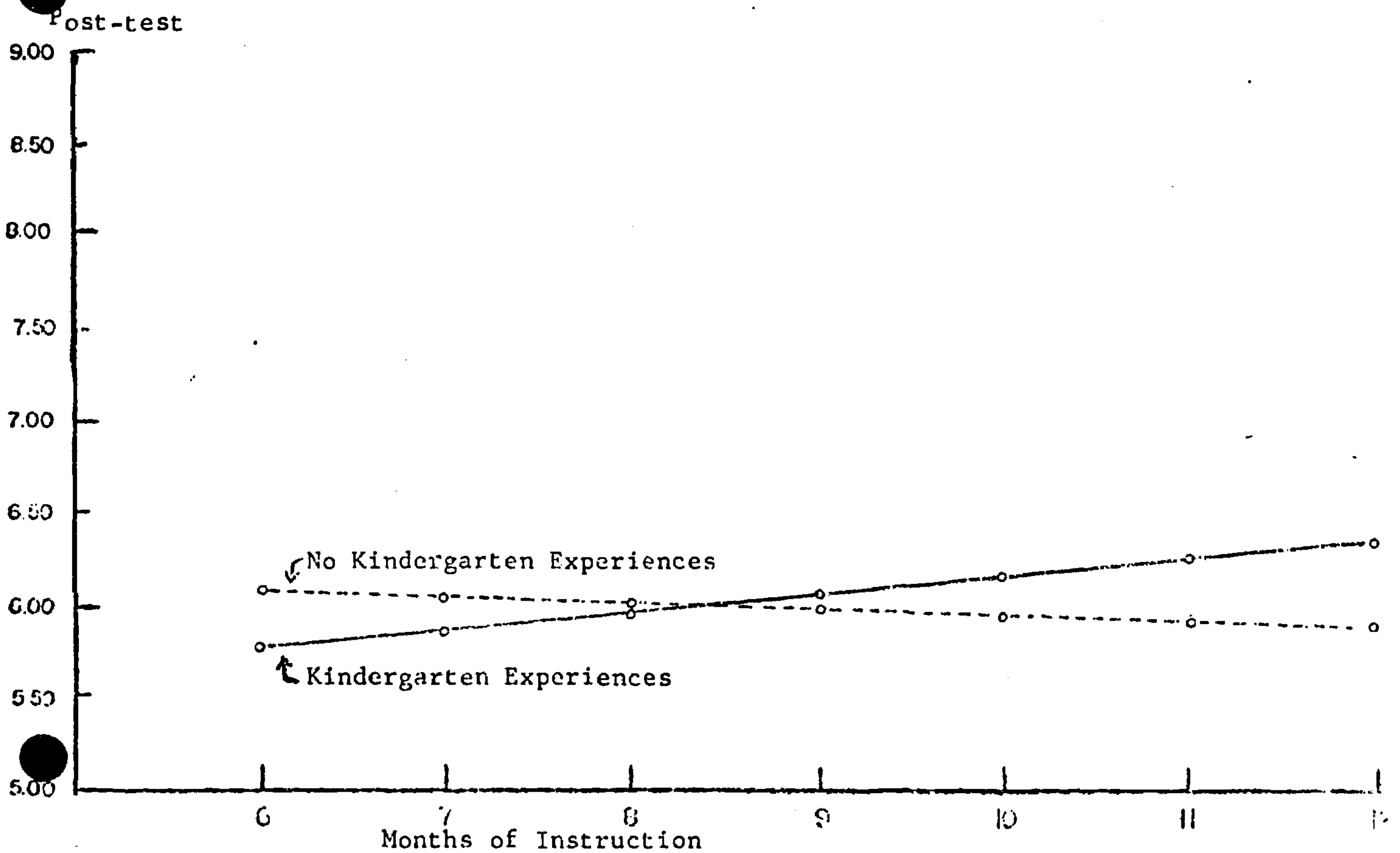

Figure 6 - Estimates of the Expected Values for Post-test Performance in reading by students scoring 5.2 on the Pre-test.

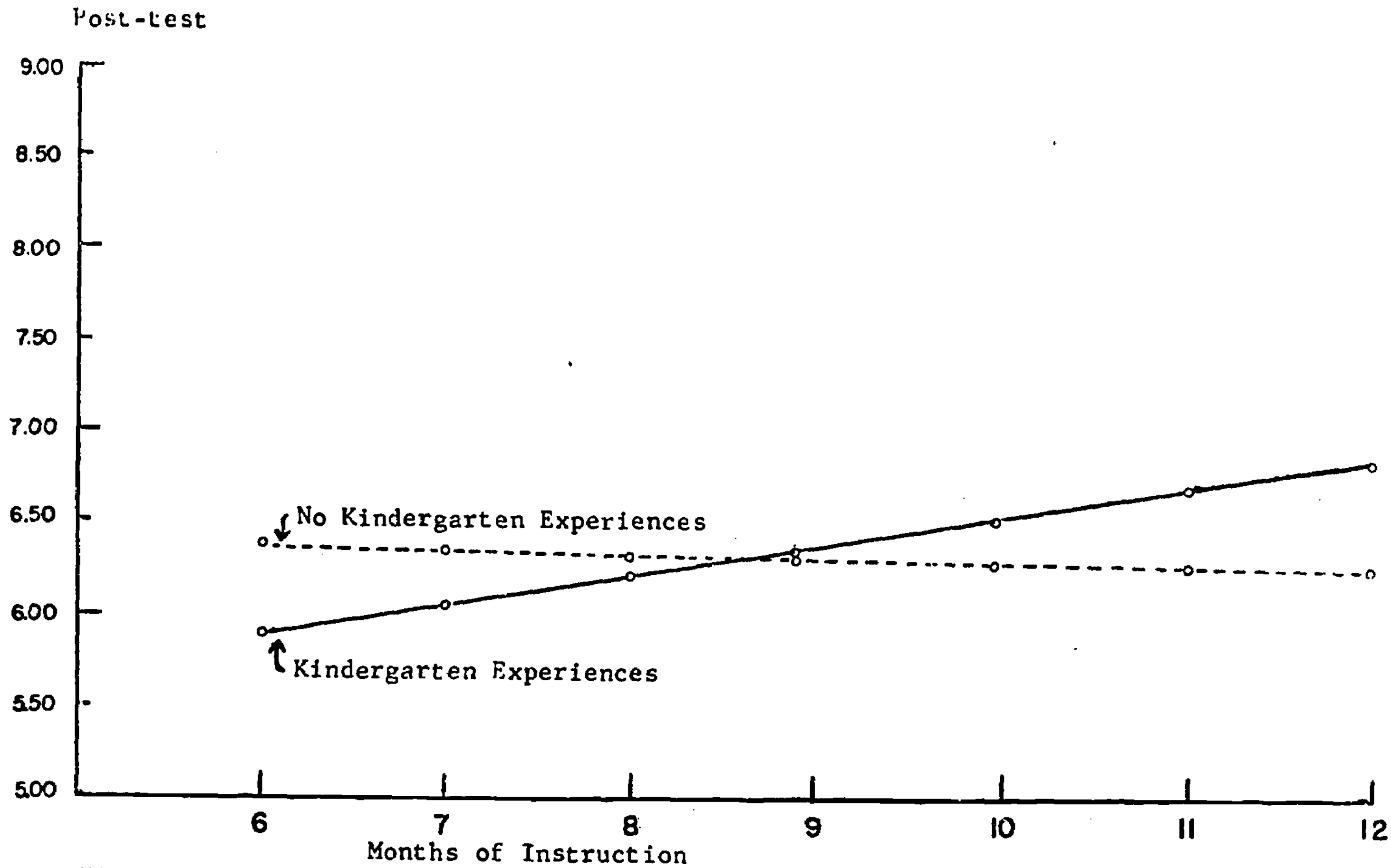

Figure 7 - Estimates of the Expected Values for Post-test Performance in reading by students scoring 5.6 on the Pre-test.

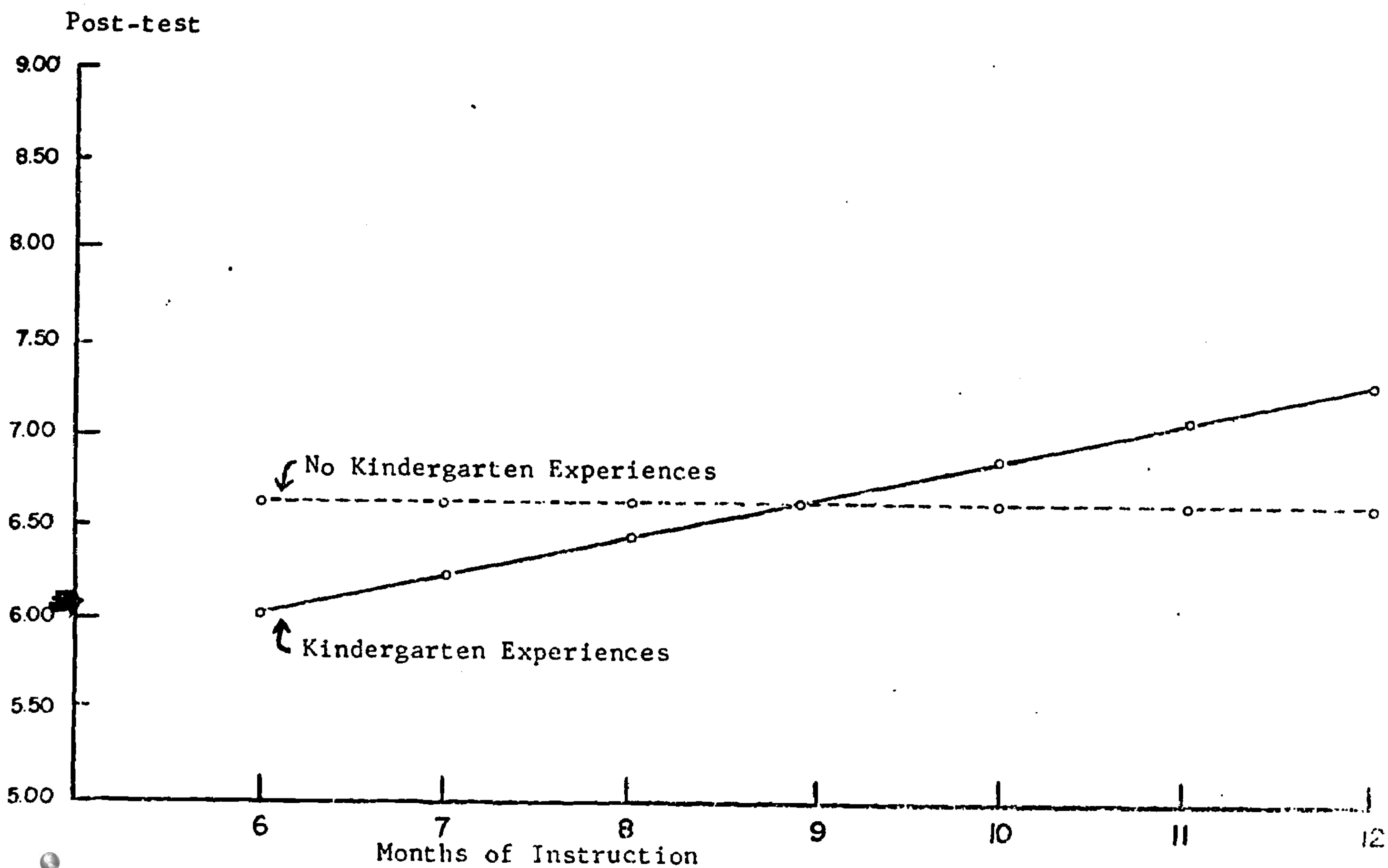

Figure 8 - Estimates of the Expected Values for Post-test Performance in reading by students scoring 6.0 on the Pre-test.

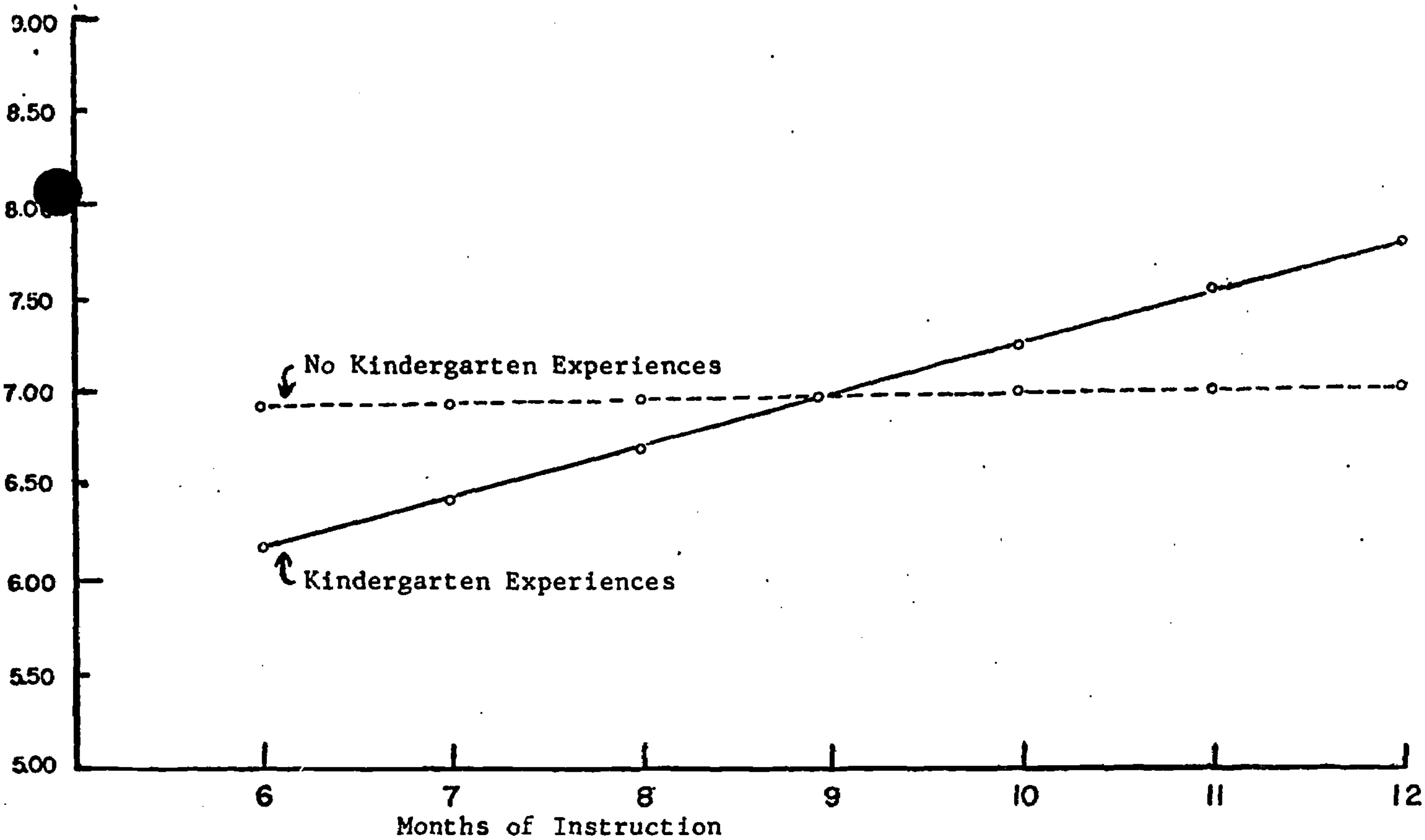

Figure 9 - Estimates of the Expected Values for Post-test Performance in reading by students scoring 6.4 on the Pre-test.

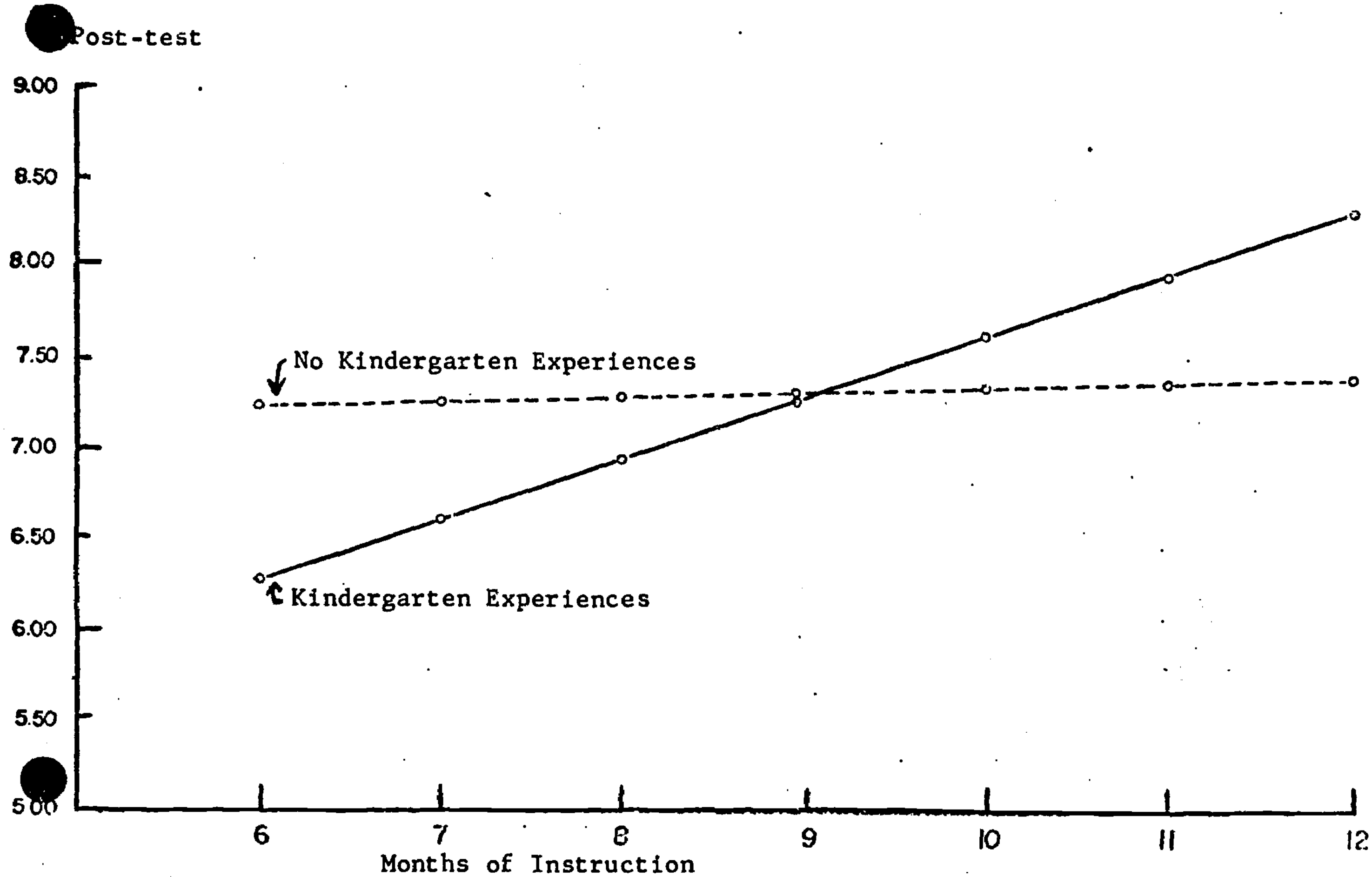

Figure 10 - Estimates of the Expected Values for Post-test Performance in reading by students scoring6.8 on the Pre-test.

GROUP	SECOND GRADE	FOURTH GRADE	SIXTH GRADE
Reading, Kindergarten	.094	.102	.079
Reading, Non-Kindergarten	.083	.094	.079

Table 2

You will note from this analysis that the second and fourth levels appear to have gained in performance as a result of the early childhood experience while this effect has been negated by level six. Although the gain score analysis takes into account the time between pre- and post-testing at these levels, the significance of that time factor is lost in the group average.

The expected value plots obtained through the linear model described, allows one to consider the effects of both the time interval factor and the differences in pre-test performance. We are now able to examine the effects of early childhood experiences upon the performance of groups of students who were comparable upon entry into the instructional process for that year and who had approximately the same amount of instruction during the year.

An examination of these plots (Figures 1-4) reveals that early childhood education has a positive impact upon the reading achievement of students at the second level but that this impact lessens over time. It may also be observed (Figures 1-4) that the groups with higher entry points require shorter periods of time before the positive effects of early childhood education are lost.

Although the gain score analysis indicated that the effects of early childhood are lost by level six, the linear model plots (Figures 5-10) at this level do not confirm this. You will note, however, that in all of the graphs presented for this level, the groups with early childhood experiences did less well in the shorter time periods. As the performance is plotted for the longer time periods, it appears that the early childhood experience is beneficial.

THE DECISION SELECTED

Today we have reviewed only a few of the 30 different analyses which were accomplished using the multiple linear regression technique. These are sufficient to illustrate the value of this technique in generating decision options for planning and evaluation.

After examining the various expected values which resulted from the linear model just described, and considering the quality of the data which was available for use with this model, we concluded that a more carefully controlled study of early childhood education is needed before we can recommend any action be taken with respect to such programs in Texas.

Early childhood education programs are being conducted in most Texas schools. We think we have gained some useful insights into better ways to evaluate these programs. It is our hope that we have provided you with some new ideas that can make your planning and evaluation effort a little better and thereby make the education available to students in your schools more meaningful.

SUMMARY

What we have tried to say can be summarized in the following statements:

- decision-making is an integral part of the process of evaluation and program planning.
- data processing in general is assuming an ever-increasing role as the vehicle for information required by decision-makers.
- two measures of our success in this role are
    1. the extent to which the decision-maker experiences freedom in phrasing questions.
    2. the degree to which our systems accurately reflect the process symbolized by the data.
- the purpose of this paper has been to show how "Multiple Linear Regression" satisfies these two measures.

BIBLIOGRAPHY

1. Allerhand, Melvin E., Head Start Operational Field Analysis, Progress Report III, Western Reserve University, Cleveland, Ohio, April 15, 1966, pp. 3-8.

2. Silberberg, Norman E., The Effects of Kindergarten Instruction in Alphabet and Numbers on First Grade Reading, Final Report. Office of Economic Opportunity, September 27, 1968, p. 14.

3. Southwest Educational Development Laboratory, Central Cities Educational Development Center, Fort Worth, Texas. Austin, Texas, June 15, 1970, pp. 91-93.

4. Southwest Educational Development Laboratory, Model Cities Early Childhood Educational Learning System of the Edgewood Independent School District of San Antonio. Austin, Texas, August 1970, p. 46.