189-73

INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

EVALUATING PROJECT HEAD START

Burt S. Barnow

UNIVERSITY OF WISCONSIN-MADISON

EVALUATING PROJECT HEAD START

by

Burt S. Barnow
University of Pittsburgh

December 1973

Financial support for this study was provided in part by the National Institute of Education, U.S. Department of Health, Education, and Welfare pursuant to Grant No. OEG-0-72-1384, and by the Institute for Research on Poverty, University of Wisconsin-Madison through funds granted by the Office of Economic Opportunity pursuant to the provisions of the Economic Opportunity Act of 1964. This paper was presented at the November 1973 meeting of the Southern Economic Association and is based largely on the author's dissertation. The reader is referred to that work for more detailed coverage of the topic. The author wishes to acknolwedge comments and suggestions from Arthur S. Goldberger, Robert H. Haveman, and especially Glen G. Cain. The opinions expressed in this paper do not necessarily reflect those of the sponsoring agencies. The author retains responsibility for any errors that remain.

ABSTRACT

Head Start is a federally-funded preschool program for disadvantaged children. In this paper the problems of evaluating Head Start from an economic viewpoint are considered and a reanalysis of the data collected for the 1969 Westinghouse Learning Corporation-Ohio University study is carried out. An evaluation of Head Start is difficult because of conceptual and statistical problems. The conceptual problems deal with the difficulty in translating the benefits of Head Start into dollar terms so that it can be determined if the benefits exceed the costs and if Head Start is the most efficient social action program. Statistical problems arise because it is rarely possible to measure all of the independent variables in the model and some of the independent variables may be measured with error; the paper discusses ways in which these specification errors can lead to biased estimates of the effects of Head Start.

Regression analysis is used in the reanalysis of the Westinghouse data. The statistical model is similar to the one used in the original study, but several modifications have been made. The findings of the reanalysis are compatable with those of the Westinghouse study but differ because of the changes in the structure of the model. The reanalysis suggests that Head Start produces statistically significant cognitive benefits for white children from mother-headed families and minority children; there is no evidence from the data to suggest that these benefits are permanent.

EVALUATING PROJECT HEAD START

1. Head Start as a Social Action Program of Interest to Economists

In recent years great efforts have been made to integrate the disadvantaged into the mainstream of the American economy. Programs have been developed to retrain workers with obsolete skills and to aid youths with less marketable skills. Economists have played a major role, both devising programs of this nature (e.g., the Job Corps) and evaluating them. While economists were emphasizing the role of training adults and teenagers already in the labor force, educators and psychologists were stressing the need for intervention at a much earlier stage in life--from the ages of three to five. Throughout the past decade many such preschool programs for disadvantaged children have been implemented, with the largest in size being the federally-funded Head Start program. In this paper how Head Start can be evaluated from an economic viewpoint is considered and the conceptual and empirical problems encountered in such an evaluation described. A reanalysis of the data collected in 1969 for the first and only national evaluation of Head Start is then presented.

Head Start is a national preschool program whose purpose has been to prepare children from disadvantaged backgrounds for entrance into formal education in the primary grades. The philosophy underlying the program is that one reason children from disadvantaged backgrounds perform poorly in school, and hence drop out and remain impoverished, is that the home environment does not provide the stimulation and amenities found in middle-class homes. By intervening between the ages of three and five, the program seeks to give these children a "head start" in their attitudes and cognitive development, and thus break out of the cycle of poverty. Although originally prepared as an experimental pilot program in 1965, Head Start was greeted with such outstanding popularity that the funding for the first year was increased from $17 million to $103 million. The program continued to grow, both in popularity and in size, and it remains as one of the few remnants of the Johnson Administration's antipoverty programs to retain widespread support.

Because Head Start has been established as a decentralized program, individual centers differ in their structure and curriculum. In addition to attempting to raise the level of cognitive development, Head Start programs have tried to improve the physical and mental health of the participants, encourage parents to take an active role in the development of the children, and improve the sense of dignity and self-worth of the Head Start participants. All of these components are important aspects of Head Start, but Head Start differs from day care in that it attempts to raise the participants' cognitive development. Because of this, and because the cognitive benefits are easier to measure than the psychological and health benefits, this paper is restricted to evaluating the cognitive benefits of Head Start.[1]

Although the evaluation of educational programs has traditionally been in the realm of psychology, in recent years economists have become interested in the field. Like on-the-job training and manpower training programs, education can be viewed as an investment in human capital; if the program is successful, the participant will be able to use his human capital to increase his earnings. Economists have generally limited their attention to the evaluation of "final" investments in human capital such as higher education and manpower training. Regardless of the effectiveness of such programs, the increasingly popular belief that preschool programs are an important instrument for helping disadvantaged children requires that efforts be made to evaluate programs such as Head Start.

There are two reasons why economists have done little research in evaluating preschool programs. First, such efforts have traditionally been carried out by psychologists. If, however, limited funds are available for social action programs, decisions must be made on how to allocate these funds between preschool children and adults in the labor force. It is the economist who is best able to determine the relative efficiencies of these two approaches. Second, it is very difficult to measure the benefits of preschool programs in economic terms. When a person participates in a manpower training program, the benefits can be measured as the difference in the present values of what the participant earns after the program minus what he would have earned had he not participated; although some problems exist in making these measurements, reasonable evaluations have taken place. For preschool education programs, however, there is a long lag between the termination of the program and the commencement of the monetary returns to the program. All that can be measured at the conclusion of the program are the benefits in terms of cognitive development. To perform a cost-benefit analysis, benefits must be converted into monetary gains that will accrue to the participants (and the rest of society) at least eight years after the completion of the program. We have obviously not reached the stage where such cost-benefit analyses can be carried out, but we can summarize the progress that has been made and note areas that will require additional research.

2. Problems in Evaluating Project Head Start

To evaluate a social action program such as Head Start it is necessary to measure both its costs and benefits. The costs can be determined in a straightforward manner, and need not concern us here. As we have noted previously, the benefits are much more difficult to ascertain. Problems in measuring the benefits of Head Start can be divided into conceptual problems and empirical problems, and each of these areas is discussed below.

The conceptual problems are those that arise in developing the theoretical model for the evaluation. The first step in the formulation of such a model is the determination of what we wish to measure as the program's output. The immediate products of Head Start may include cognitive and psychological benefits of the program, day-care benefits to the parents of the participants, and medical and nutritional benefits to the children. These benefits may, however (with the possible exception of the day-care benefits), be considered only intermediate goods which serve as inputs in the production of the output of ultimate interest to economists--earning ability.

For an evaluation of Head Start we can divide the procedure into four steps: (1) the determination of our best estimate of the effect of Head Start on the cognitive development of the participants for various groups of children; (2) the consideration of how confident we can be in the accuracy of the estimates; (3) the determination of whether the benefits of the program exceed the costs; and, (4) a comparison of Head Start to policy alternatives to determine if Head Start is the most efficient program for meeting our goals. If the first and second steps indicate a zero or negative effect of Head Start, the subsequent steps are unnecessary and Head Start can be viewed as an ineffective program.

To estimate the effects of Head Start, a model must be specified that includes Head Start as one of the inputs. Because the dependent variable in the model is the cognitive development of the child, the model can be viewed as an educational production function where we attempt to relate the various input factors supplied by the child, his family and society to the output of education. Research concerning educational production functions is still in the early stages, and there are problems in determining which variables should be included and what functional form should be used. To simplify the discussion, it is assumed that a linear functional form is appropriate (at least for the range of the available data), but it should be noted that substantial research remains to be done in this area. Head Start can be considered as either a binary variable (where a child either participates in the program or does not) or as a continuous variable (where the length of the treatment can vary by increasing either the hours per day spent in Head Start or the number of weeks that the program lasts). For this illustration it is assumed that Head Start is a continuous variable. Data must then be collected on the variables in the educational production function for a group of children. For reasons discussed below, the nature of the data collected will be important in determining if the data will produce unbiased estimates of the coefficients in the educational production function. The dependent variable in the model is some measure of cognitive development taken after the completion of the Head Start program. Assuming that the model has been properly specified, the cognitive measure may be regressed on the set of independent variables, and the regression coefficient for the Head Start variable will be the best estimate of the effect of one unit of Head Start on the cognitive measure used as the dependent variable. If the coefficient of Head Start is zero or negative, our work is done--it may be concluded that in its present form Head Start is an inappropriate program for increasing the cognitive development of the participants. However, if the effect of Head Start is positive, the statistical significance of the coefficient must be considered. If the hypothesis of a zero effect cannot be rejected at the 1 or 5 percent significance level it is sometimes argued that the program should be abandoned because there is no "significant" effect. This argument is rarely correct. As Cain and Watts (1970, p. 233) explain, "a body of data may be unable to reject the hypothesis that some coefficient is zero and be equally consistent with a hypothesis embodying a miraculously high effect." It must be kept in mind that the regression coefficient is the best estimate of the treatment effect and that a large standard error for the coefficient only implies that we cannot be very confident in the accuracy of the estimate. If the regression indicates the Head Start coefficient is not significant but that the coefficient is large enough so that the effect of the program appears to be worth attaining, a more refined evaluation of Head Start should be attempted. Some of the techniques that can be used to increase the efficiency of the estimate include: increasing the sample size, improving the accuracy of the measurement of variables in the model, and selecting the sample so that there is less covariation between Head Start and the other independent variables. For policy decisions the choice of when to dismiss a coefficient as being insignificant should be based on the costs of making an incorrect decision rather than on some arbitrarily selected significance level.

A Head Start instructor may be satisfied to learn only what the cognitive benefits of Head Start are, but a policymaker must know more. The policymaker must know whether the benefits of Head Start exceed the costs of the program so he can determine if the program is worth funding. The relevant question is not "Does the program work?" but rather, "Is the program worth funding?" This part of the evaluation requires placing a dollar value on cognitive gains--a very difficult task, but one which must be faced. If it costs $1,500 to make Head Start available to one child, then the policymaker must decide if the gains produced by Head Start are worth at least that much for the program to be judged worthwhile. For private goods, economic theory asserts that an individual will purchase a good only if the benefits he receives exceed the costs. For publicly funded social action programs, however, it becomes much more difficult to price the benefits, especially when the benefits are in the form of gains in cognitive development for young children. The benefits of the program would not necessarily be the discounted increase in earning power due to the change in IQ that occurred because of Head Start. If the benefits of Head Start fade after the program, then such a procedure would overstate the value of the gains. It is also possible that the gains from Head Start would allow the child to be placed in a higher "track" in the primary grades, thus leading to additional gains. Finally, the externality benefits to the taxpaying public and the day-care benefits to the child's parents should be included in the benefits.

After the benefits of Head Start are expressed in dollar terms, the benefits can be compared with the costs of the program. If the costs exceed the benefits, then Head Start should not be continued in its present form even if it produces significant benefits. For example, we would reject a manpower training program that increased the percent value of an individual's lifetime earnings by $10,000 if the program cost $15,000 per participant.

The final step in the evaluation is to compare Head Start with alternative policies. To illustrate this point we shall use the following hypothetical example: assume that the educational production function included three policy variables: Head Start (Z), kindergarten (K), and cash transfers (I). Further assume that Head Start and kindergarten are continuous variables. Children can receive any combination of the policy variables, but it is assumed that the relationship between the cognitive test score (Y) and all independent variables is linear and additive with no interaction effects.

The first step in the analysis is to regress the cognitive measure Y on the three policy variables and all other independent variables specified by the model. Suppose that the fitted regression equation is:

$$Y = \beta_o + .0005I + 1.00Z + 5.00K + \sum_{i=1}^{N} \beta_i X_i$$

where the X variables represent the nonpolicy variables such as age, race, and sex. Because the regression coefficient for Head Start is positive, we can proceed to compare the costs and benefits of the program. For this example assume that one unit of Head Start costs $1,000 per child. If the policymaker decides that a gain of one point on the dependent variable and the other gains from the program are worth at least $1,000 then the program can be judged a success. The policymaker can then compare Head Start with the policy alternatives to see which is the most effective method of increasing the cognitive development of disdvantaged children. Suppose that the following cost information is available:

cost of a unit of Head Start = \$1,000

cost of a unit of kindergarten = \$2,000

cost of income transfer = \$1

By combining the cost information with the regression coefficients we can arrive at the cost-benefit ratios for the three policies:

points/dollar for Head Start = .0010

points/dollar for kindergarten = .0025

points/dollar for income transfer = .0005

Thus for the example presented here, Head Start is not the most efficient way of increasing the cognitive development of children; kindergarten provides more benefits per dollar of expenditure.[2]

It is considerably more difficult to compare Head Start with social action programs for those already in the labor force. For such comparisons it is necessary to either select some tradeoff between cognitive gains for preschool children and earning power for adults or extrapolate the cognitive gains and determine how they affect earning power. As most evaluations of preschool programs have found that the gains disappear within three years, using the second criteria would lead to the abandonment of preschool programs as a means of increasing earning power.

The procedure for determining the appropriate policy instrument becomes more complex if externalities are considered or if Head Start and kindergarten can be provided at one treatment level. When externalities are present, the program that is most efficient for increasing cognitive development or earning power may not be as politically acceptable as an alternative policy. Thus, even if a manpower training program were more effective than Head Start, taxpayers might prefer that their money be spent on children rather than on adults. Indeed, the only justifications for in-kind transfers are paternalism and certain types of externalities. If Head Start and kindergarten are available only in one quantity (where a child either participates in a program or he does not, but the duration of the program is fixed) then there is an interesting equity-efficiency tradeoff. In the example presented above, the largest aggregate amount of benefits would be produced by concentrating all of the funds on kindergarten; yet this procedure will provide benefits to a smaller number of children and may be criticized on equity grounds. The policymaker must then decide how to weight the equity and efficiency aspects in making his decision.

The formulation of a conceptual model offers no guarantee that an unbiased evaluation can be carried out. For evaluations of programs such as Head Start it is generally not possible to directly test the theoretical model. Specification errors occur because one or more of the independent variables are undesirable or are measured with error.

To illustrate these problems several simple models will be developed; a more complete discussion of these models and several additional models can be found in the works by Barnow (1972 and 1973) and Goldberger (1972a and 1972b). The underlying model considered is one where a child's cognitive development in period 2 is a linear function of his cognitive development in period 1 and the gains he receives from participating in Head Start. To simplify the analysis Head Start is assumed to be available at only one level, and the dummy variable in Head Start, Z, is defined so that Z=1 if a child has participated in Head Start and Z=0 if he has not. Cognitive development in period 1, prior to Head Start experience, is represented in the model as X^*, and measured cognitive development in period 2, after Head Start experience, is denoted as Y.

It is important to note that the variable Y is a fallible measure of cognitive development but that X* contains no measurement error. (In the psychological literature X* is referred to as a true score.) To complete the model, the disturbance term associated with the posttest is denoted as v. The model may now be written as:

$$Y = \beta_o + \beta_1 X^* + \beta_Z Z + v. \tag{1}$$

It is necessary to include a control group of children who have not participated in Head Start to estimate β_Z, which is the treatment effect. If we simply regressed Y on X* for children who have had Head Start, we would be unable to differentiate between normal cognitive growth and the effect of Head Start; this problem would be especially great if the treatment is of long duration.

Unfortunately it is rarely possible to directly estimate equation (1). The level of cognitive development prior to the treatment is usually unavailable or available with measurement error. Tests that can be used to measure pretreatment cognitive development may provide an unbiased measure of X*, but they are subject to error. Denoting a fallible but unbiased measure of pretreatment cognitive development as X we may write

$$X = X^* + u \tag{2}$$

where u is assumed to be normally distributed with a zero mean and uncorrelated with X*. The variable X is generally referred to as a pretest and Y is referred to as a posttest. It will also be assumed that u is independent of v, X*, and Z; more formally this can be stated as:

$$\text{Cov}(u,v) = \text{Cov}(u,X^*) = \text{Cov}(u,Z) = \text{Cov}(v,X^*) = \text{Cov}(v,Z) = 0 \quad (3)$$

Assuming that X and not X* is available for an evaluation, we are interested if the regression coefficient of Z will be the same when the linear regression of Y on X and Z is run rather than Y on X* and Z.[3] Thus when we determine

$$E(Y|X,Z) = \alpha_0 + \alpha_1 X + \alpha_Z Z \quad (4)$$

we wish to discover if the α coefficients are equal to the corresponding β's. It can easily be demonstrated that, in general, errors in the measurement of one of the regressors leads to bias of all regression coefficients. For an evaluation of Head Start it is especially important to learn the relationship between α_Z and β_Z. First we make the following definitions:

$$\sigma_{**} = \text{Var}(X^*),\ \sigma_{ZZ} = \text{Var}(Z),\ \sigma_{*Z} = \text{Cov}(X^*,Z),\ \sigma_{uu} = \text{Var}(u), \quad (5)$$
$$P = \frac{\sigma_{**}}{\sigma_{**}+\sigma_{uu}},\ r^2 = \frac{(\sigma_{*Z})^2}{\sigma_{**}\sigma_{ZZ}}\ .$$

Note that the parameter P is the ratio of the variance of the true measure of pretreatment cognitive development to the variance of the pretest and that $0 \le P \le 1$. The parameter r^2 is the squared coefficient of correlation between X* and Z. When the normal equations for (4) are solved, it can be shown that:

$$\alpha_Z = \beta_Z + \frac{\sigma_{*Z}(1-P)\beta_1}{\sigma_{ZZ}(1-Pr^2)} \quad (6)$$

There is no bias in equation (6) only under certain conditions. One example is if P=1 then $\alpha_Z=\beta_Z$, but this requires that X*=X and that there is no measurement error. Another is when $\beta_1=0$; i.e., the variable measured with error does not belong in the regression. This case is theoretically untenable because it implies that a child's cognitive development in the second period has no relationship to his cognitive development in the first period. A more interesting case is that $\sigma_{*Z}=0$ suffices for $\alpha_Z=\beta_Z$. But $\sigma_{*Z}=0$ is equivalent to $E(X^*|Z)=E(X^*)$, which says that the mean of pretreatment cognitive development is the same in the experimental and control groups. Random selection and matching on pretests would lead to this result and therefore avoid bias. Note that if $\sigma_{*Z}<0$ (the experimental group initially has lower cognitive development) then $\alpha_Z<\beta_Z$ and the bias would be to underestimate the treatment effect.

The primary lesson to be learned from this exercise is that when one of the regressors is measured with error the estimate of the treatment effect may be biased. If the Head Start administrators used a "scraping" procedure whereby children with the lowest levels of true cognitive development were assigned to the experimental group and the central group selected from the more able children, regression analysis would lead to an underestimate of the true effect of Head Start. Random assignment to the experimental and control groups is the most appealing way to avoid bias in evaluations, but random assignment is not always possible. In many instances an evaluation must be carried out ex post facto or certain criteria must be used for admitting children to the experimental group.

The data which we analyze in the next section of this paper were collected for an ex post facto evaluation of Head Start. Although no direct measure of pretreatment cognitive development is available, when certain assumptions are made the socioeconomic and demographic data collected at the time of the posttest can be used as a proxy of pretreatment cognitive development. The following are the assumptions that must be made to utilize the socioeconomic and demographic variables: (1) the variables must be a function of cognitive development, (2) we must know how to specify the relationship between the set of independent variables available and cognitive development, and (3) exposure to Head Start does not affect the socioeconomic status of a child. If we make these assumptions then the information collected on the background of the Head Start and control children can be used as a fallible measure of pretreatment cognitive development in statistical analyses.

An unfortunate drawback of ex post facto analyses is that unless the process by which children were assigned to the treatment groups is known, it is impossible to determine if regression analysis will produce unbiased estimates of treatment effect; it may even be impossible to discuss the direction of the bias. For the errors-in-variables model developed above, discriminant analysis can be used to test the hypothesis that the Head Start and control groups differ significantly on a set of variables that is associated with cognitive development. However, that model is consistent only with certain selection procedures--such as when selection is based on true scores or when selection is based on one set of variables and another set is available for the evaluation--and other selection procedures may not lead to biased estimates when the Head Start and control groups differ in their pretreatment cognitive development.

For example, if all eligible children were ranked on some socioeconomic measure and the children in the lower half of the distribution were entered in Head Start while the upper half were then placed in the control group, the control group would have a higher mean level of pretreatment cognitive development than the Head Start group. But it has been demonstrated by

Goldberger (1972a) and Barnow (1972) that this selection procedure will not lead to biased estimates of the Head Start effects when the posttest score is regressed on the selection variable (X) and the dummy variable for Head Start. The reason that the errors-in-variables model described above does not apply when this selection procedure is used is that the assumption that Cov(u,Z)=0 is violated. Several additional selection procedures are analyzed by Barnow (1973) and Goldberger (1972a), and it is demonstrated that one must know the selection procedure used before any conclusion can be drawn on whether a given data set will produce biased estimates of the effects of Head Start.

3. Estimating the Cognitive Benefits of Head Start

By 1969, Head Start had been in existence for three years, and officials at the Office of Economic Opportunity felt that the program was ready to be evaluated. Several long-term evaluations were planned including a seven-year longitudinal study and a planned variation study. However, these evaluations would not provide useful information for several years so a contract was awarded to the Westinghouse Learning Corporation in conjunction with Ohio University to carry out a one-year ex post facto evaluation of summer and full-year Head Start programs throughout the nation. Although the Westinghouse researchers attempted to determine the affective as well as the cognitive benefits of Head Start, in this paper attention is restricted to the latter because I feel that not only are cognitive aspects more important but there is little evidence to indicate that the affective tests designed for the Westinghouse evaluation are satisfactory.[4]

To carry out the study the Westinghouse researchers randomly selected 300 Head Start centers from a list of all centers. From this sample 225 were screened until 104 centers (29 full-year centers and 75 summer centers) were found that could participate in the evaluation; although Smith and Bissell (1970) argue that the dropped centers were the most effective ones, they offer no strong evidence to support this view. From each center a random sample of eight former Head Start participants who were in each of the first, second, and third grades at the time of the evaluation were selected. The Westinghouse researchers recognized the weaknesses of ex post facto evaluations and established the following criteria for considering children for the Head Start or control populations:

> (1) Continuity of residence in the target area. All subjects must have lived in the target area from the time of the specified Head Start program until the time of the study...
>
> (2) Eligibility for Head Start. All subjects must have met the eligibility requirements for participating in the Head Start program.
>
> (3) Equivalent school experience. All must have attended the same school system.
>
> (4) No other Head Start experience. (Cicirelli et al., p. 36. For the remainder of the paper all page references refer to Cicirelli et al., unless otherwise specified.)

Within each of the 104 neighborhoods a control sample of eight children at each grade level was then selected. In addition to fulfilling the four criteria listed above, the control children were selected to match the Head Start sample on the basis of age, sex, kindergarten attendance, and social/ethnic characteristics. The groups were not matched in socioeconomic status (SES), but information on SES was collected and used as a covariate in the statistical analysis. Information on the SES of the children was collected by interviewing their parents. In addition, children in the final sample were given a series of cognitive and affective tests.

The primary method of statistical analysis used by the Westinghouse researchers was analysis of covariance. The equivalent regression equation for their analysis is

$$Y = \beta_o + \beta_Z Z + \beta_1 X_1 + \sum_{j=2}^{N} \beta_j X_j + u$$

where the variables are defined as:

Y	= the score on the cognitive test
Z	= a dummy variable for experimental status where $Z = 1$ for children who participated in Head Start and $Z = 0$ for the children in the control group
X_1	= the measure of SES used in the study
$X_2 \ldots X_N$	= a series of dummy variables for the N neighborhoods or target areas where $X_j = 1$ if the child is from neigborhood j and $X_j = 0$ otherwise
u	= a disturbance term assumed to be normally distributed with a mean of zero and constant variance, and assumed to be uncorrelated with the independent variables.

The Westinghouse researchers felt that the data should be grouped by neighborhood rather than use individual children as the unit of observation; for each neighborhood there were thus two observations--one for the Head Start children and one for the control children. The mean value for each group on each variable was determined and used for the values for that observation.

The variable used to control for SES in the Westinghouse study is the "Two-Factor Index of Social Position" described in Hollingshead (1958). The Hollingshead Index is a weighted sum of the head of household's occupation and educational attainment where both factors have been divided into a seven-point scale.

The children in the samples were given both an ability (intelligence) test and an achievement test. The ability test used in the study is the revised edition of the Illinois Test of Psycholinguistic Abilities (ITPA). The object of the ITPA, according to Kirk et al. (1970, p. 5) is to "delineate specific abilities and disabilities in children in order that remediation be undertaken when needed. It is a diagnostic test of specific cognitive abilities as well as a molar test of intelligence." The ITPA is structured so that it is applicable for children from the ages of two to ten, and it was administered to the children in all three grades. The achievement tests that the children took are the Metropolitan Readiness Test, Stanford Achievement Test Primary I Battery, and the Stanford Achievement Test Primary II Battery for the first, second, and third grades, respectively.

Each of the cognitive test scores was used as a dependent variable for full-year and summer samples for the first, second, and third grades. In addition, the samples were further stratified by geographic region, racial/ethnic characteristics of the centers, and by type of population units in which the centers were located. Most of the analyses conducted by the Westinghouse researchers indicated a zero or negative effect for summer Head Start and a small positive effect for full-year Head Start. The study concludes that:

> In summary, when one looks at the observed effects of Head Start according to the test of practical relevance, it must be concluded that the effects found on the standardized tests are indeed small in magnitude, with the exception of a few differences found in subgroups of full-year centers on the ITPA, and do not meet the criterion of practical relevance (p. 168).

The Westinghouse report concludes that summer Head Start programs should be phased out as soon as possible and replaced by full-year programs, and that attempts should be made to strengthen the full-year programs.

Perhaps because of the ex post facto design, and to some degree because of the unpopular nature of the findings, the Westinghouse study has been the subject of many criticisms. Although some critics have argued that the questions addressed by the study are incorrect and that the cognitive tests used may be inappropriate, the most serious criticisms deal with the selection of the Head Start and control children for the study and whether or not the estimate of the effects of Head Start found in the study are biased. Although the crucial question of whether or not the Westinghouse data can be used to produce unbiased estimates of the effects of Head Start can never be answered, some critics feel that the Westinghouse researchers did not analyze the data in the best way and that a reanalysis would produce a better estimate; indeed, the results of our reanalysis are consistent with the Westinghouse findings but demonstrate that Head Start may be effective for certain groups of children.

To improve the statistical analysis of the Westinghouse data, we have made several changes in the regression equations used. Whatever the merits of the Hollingshead Index are in measuring status, there is no reason for including socioeconomic information in that form for a model of cognitive development. Although the information that is used in computing the Hollingshead Index can be included in the model either to serve as a proxy for pretreatment cognitive development or as independent variables in their own right, the information can better be included as sets of dummy variables for two reasons: (1) the Hollingshead Index constrains the coefficient for educational achievement to be four-sevenths of the coefficient for occupational status, whereas including the variables independently permits the coefficients to vary freely;[5] and (2) when the information is entered as precoded variables the effect of moving from any one step to the next highest is constrained to be the same, but the use of dummy variables allows the effect to vary from step to step. As knowledge concerning the relationship between cognitive development and SES is very limited, one should permit the functional form to be as general as possible.

Variables not included in the Hollingshead Index are also appropriate for inclusion in the regression model. By expanding the set of independent variables we expect to reduce the possible bias in the coefficient for Head Start caused by nonrandom treatment assignment. In addition, the variables that we have included are useful in helping us to learn more about the educational production process, and several of the variables can be used as policy variables in addition to or in place of Head Start. A list of the variables employed in the reanalysis and a description of how the variables were formed is given in Table 1. When a single trait, such as mother's education, is represented by a group of dummy variables (i.e., MSOCOL, MHSG, MSOHS, M79, and M06) we have generally followed the practice of omitting the variable for the highest category from the regressions.

A second change that has been made for the reanalysis is that individual rather than grouping data are used. Cramer (1964) has demonstrated that grouping leads to a loss of efficiency, and Blalock (1961, pp. 102-112) shows that grouping can sometimes bias the regression coefficients. For these reasons we have used individual data and thereby increased the number of observations by a factor of eight.

The final change in the method of analysis that we have employed is that we have stratified the data differently than the Westinghouse researchers did to permit Head Start to have a different effect for various classes of

TABLE 1

Description of Variables Used in the Reanalysis

Variable	Description
CHILD	Number of children in the family
INCOME	Total annual income of the child's family in dollars
AGE	Age of the child to the nearest year
MSOCOL	1 if child's mother has more than 12 years of education; 0 otherwise
MHSG	1 if child's mother has 12 years of education; 0 otherwise
MSOHS	1 if child's mother has 10-11 years of education; 0 otherwise
M79	1 if child's mother has 7-9 years of education; 0 otherwise
M06	1 if child's mother has 0-6 years of education; 0 otherwise
MOPRO	1 if child's mother has professional or managerial occupation; 0 otherwise
MOCLER	1 if child's mother has clerical occupation; 0 otherwise
MOSKIL	1 if child's mother has skilled occupation; 0 otherwise
MOSEMI	1 if child's mother has semiskilled occupation; 0 otherwise
MOUNSK	1 if child's mother has unskilled occupation or no occupation; 0 otherwise
FEMALE	1 if child is female; 0 otherwise
MALE	1 if child is male; 0 otherwise
RURAL	1 if child lives in a rural area; 0 otherwise
KIND	1 if child attended kindergarten; 0 otherwise
NOKIND	1 if child did not attend kindergarten; 0 otherwise
FSOCOL	1 if child's father has more than 12 years of education; 0 otherwise
FHSG	1 if child's father has 12 years of education; 0 otherwise
FSOHS	1 if child's father has 10-11 years of education; 0 otherwise

TABLE 1 (continued)

Variable	Description
F79	1 if child's father has 7-9 years of education; 0 otherwise
F06	1 if child's father has 0-6 years of education; 0 otherwise
FAPRO	1 if child's father has professional or managerial occupation; 0 otherwise
FACLER	1 if child's father has clerical occupation; 0 otherwise
FASKIL	1 if child's father has skilled occupation; 0 otherwise
FASEMI	1 if child's father has semiskilled occupation; 0 otherwise
FAUNSK	1 if child's father has unskilled occupation; 0 otherwise
WHITE	1 if child is white; 0 otherwise
BLACK	1 if child is black; 0 otherwise
MEXAM	1 if child is Mexican-American; 0 otherwise
HDSTRT	1 if child has had Head Start; 0 otherwise
BLKHS	1 if child is black and has had Head Start; 0 otherwise
ITPAMN	mean of child's nonzero scores on ITPA
MRTMN	mean of child's nonzero scores on MRT
SAT2MN	mean of child's nonzero scores on SAT2
SAT3MN	mean of child's nonzero scores on SAT3
DIVOR	1 if child's parents divorced; 0 otherwise
SEPAR	1 if child's parents separated; 0 otherwise
WIDOW	1 if child's mother is a widow; 0 otherwise
NEVMAR	1 if child's mother never married; 0 otherwise

children. We have stratified by type of program (full-year and summer), grade (first, second, and third), and by parents present (both and mother only). The stratification by parents present is necessary because the set of independent variables is different when only the mother is present; the variables for occupation and education of the father must be dropped and we have added variables for the marital status of the mother. There are ten rather than twelve samples because there were too few third grade full-year children to analyze. Stratification by race was attempted, but it was discovered that only the coefficients for Head Start and kindergarten differed significantly for Blacks and whites so interaction variables were added for the final analysis instead of stratifying by race.

The complete set of regression results from the reanalysis can be found in Barnow (1973). In the present paper some of the important results are reproduced and the major findings are summarized. In Table 2 and Table 3 the means and standard deviations are presented for the first grade, summer and full-year, both parents present samples. The tables indicate that the children in the sample are indeed from disadvantaged backgrounds. The average number of children per family is as high as 4.8 and annual family income is as low as $4,861. Although both the Head Start and control children come from disadvantaged families, the Head Start children are on average more disadvantaged as is indicated by income and family size. Note also that the means of the cognitive tests (ITPAMN and MRTMN) are higher for the control groups than for the Head Start groups. Thus, if we regressed cognitive development on treatment status alone we would find a negative effect for Head Start.

TABLE 2

Means and Standard Deviation for Grade 1,
Summer, Both Parents Present Sample

	Head Start		Control	
Variable	Mean	Standard Deviation	Mean	Standard Deviation
CHILD	4.69	2.10	4.12	2.04
INCOME	5049.	2514.	5859.	2986.
AGE	5.89	.446	5.95	.548
MHSG	.263	.441	.416	.494
MSOHS	.362	.481	.237	.426
M79	.194	.396	.192	.394
M06	.092	.290	.065	.247
MOCLER	.109	.137	.042	.201
MOSKIL	.032	.176	.023	.149
MOSEMI	.130	.337	.078	.268
MOUNSK	.794	.405	.834	.372
FEMALE	.511	.501	.490	.501
RURAL	.263	.441	.234	.424
FHSG	.225	.419	.295	.457
FSOHS	.222	.416	.286	.452
F79	.263	.441	.175	.381
F06	.165	.372	.123	.329
FACLER	.041	.199	.045	.209
FASKIL	.197	.398	.240	.428
FASEMI	.311	.464	.351	.478
FAUNSK	.403	.491	.263	.441

TABLE 2 (continued)

	Head Start		Control	
Variable	Mean	Standard Deviation	Mean	Standard Deviation
BLACK	.289	.454	.247	.432
MEXAM	.133	.340	.107	.310
KIND	.625	.485	.614	.488
BLKIND	.194	.396	.175	.381
ITPAMN	19.13	3.52	19.30	3.74
MRTMN	9.03	2.60	9.31	2.74
N	315		308	

TABLE 3

Means and Standard Deviations for Grade 1, Full-Year, Both Parents Present Sample

	Head Start		Control	
Variable	Mean	Standard Deviation	Mean	Standard Deviation
CHILD	4.83	2.17	4.74	2.01
INCOME	4861.	2252.	5490.	2656.
AGE	5.97	.492	5.96	.420
MHSG	.385	.489	.304	.462
MSOHS	.308	.464	.294	.458
M79	.192	.396	.196	.399
M06	.067	.252	.088	.285
MOCLER	.019	.138	.039	.195
MOSKIL	.019	.138	.029	.170
MOSEMI	.173	.380	.157	.365
MOUNSK	.769	.423	.735	.443
FEMALE	.510	.502	.539	.501
RURAL	.212	.410	.245	.432
FHSG	.308	.464	.304	.462
FSOHS	.260	.441	.196	.399
F79	.173	.380	.147	.356
F06	.192	.396	.157	.365
FACLER	0.00	0.00	.029	.170
FASKIL	.240	.429	.284	.453
FASEMI	.356	.481	.363	.483
FAUNSK	.337	.475	.245	.432

TABLE 3 (continued)

Variable	Head Start Mean	Head Start Standard Deviation	Control Mean	Control Standard Deviation
BLACK	.500	.502	.480	.502
MEXAM	.192	.396	.157	.365
KIND	.558	.499	.588	.495
BLKIND	.288	.455	.304	.462
ITPAMN	18.84	3.84	19.66	4.26
MRTMN	8.74	2.53	8.81	2.68
N	104		102	

The regression equations for the first grade, summer and full-year, both parents present samples with the ITPA score as the dependent variable are in Table 4. Some of the coefficients did not have the expected sign and require some explanation. It was expected that income would have a positive coefficient, but for the summer sample and some of the other samples, income had a negative but insignificant coefficient. This could be because the education and occupation variables capture the permanent income status of the child's family and that measured income includes transfer payments received by the family; unfortunately, the Westinghouse data do not break down the income information by source.

The information for the occupation and education of the parents is included in the regression equations as sets of dummy variables with the highest categories used as the reference group. For a nationally representative sample we would expect the coefficients within each set to become increasingly negative as we progress to lower levels of education and occupation. However, the Westinghouse data are not representative of all families. Only children who were eligible for Head Start are in the sample, and because Head Start sought to enroll disadvantaged children, the parents with high levels of education (some college) and occupation (managers and professionals) must be atypical of the general population of these parents; parents with such high levels of SES whose children are still eligible for Head Start may be considered as "failures." The implication of this sample truncation is that children whose parents are in the base groups for the educational and occupational variables may not only be expected to score lower on cognitive tests than children of parents with similar levels of attainment in the general population, but may even

TABLE 4

Effects of Individual Characteristics for Grade 1, Both Parents Present Sample, on Child's ITPA Score, Summer and Full Year

Independent Variable	Summer	Full Year
CHILD	-.134* (-1.900)	-.355** (-2.720)
INCOME	-.00001 (-.17000)	.0004** (3.2900)
AGE	.766** (2.790)	1.44** (2.52)
MHSG	-.194 (-.350)	-.291 (-.260)
MSOHS	-.508 (-.870)	-.552 (-.480)
M79	-.736 (-1.150)	-1.22 (-.97)
M06	-1.62** (-1.97)	-.512 (-.330)
MOCLER	2.42** (1.97)	-2.95 (-1.39)
MOSKIL	1.88 (1.51)	-.901 (-.400)
MOSEMI	.158 (.150)	-1.13 (-.70)
MOUNSK	.506 (.520)	-1.80 (-1.16)
FEMALE	-.120 (-.430)	.384 (.750)
RURAL	.240 (.700)	1.17 (1.03)
FHSG	.341 (.690)	-1.73* (-1.90)
FSOHS	-.003 (-.010)	-.950 (-.990)
F79	-.475 (-.890)	-2.16* (-1.99)

TABLE 4 (continued)

Independent Variable	Summer	Full Year
F06	-.905 (-1.450)	-1.07 (-.95)
FACLER	-.921 (-1.080)	1.25 (.52)
FASKIL	-1.01 (-1.63)	.576 (.490)
FASEMI	-1.17* (-1.90)	-.221 (-.190)
FAUNSK	-1.64** (-2.57)	-.040 (-.030)
BLACK	-2.10** (-3.11)	.810 (.560)
MEXAM	-1.12* (-1.69)	-.469 (-.330)
KIND	.862** (2.500)	3.47** (3.94)
BLKIND	-.593 (-.900)	-3.56** (-3.11)
HDSTRT	-.361 (-1.000)	-1.22 (-1.38)
BLCKHS	1.99** (3.13)	2.04* (1.79)
MEXHS	1.53* (1.78)	.081 (.050)
CONSTANT	16.67** (1.98)	12.00** (2.82)
R^2	.186	.374
N	623	206

Note: t-statistics are in parentheses below their coefficients.

*Statistically significant at the 10 percent level.

**Statistically significant at the 5 percent level.

score lower than children in the Westinghouse sample whose parents are in some of the lower categories.

Black and Mexican-American children have been observed to score lower than white children on many ability and achievement tests. We lack the expertise to determine if such findings are due to real differences or cultural bias in the tests. It should be noted that for the Westinghouse data the Black and Mexican-American children come from significantly more disadvantaged backgrounds than the white children, and the negative coefficients found for minority group children may be due in part from failure to adequately control for these differences.

Head Start and kindergarten are the two most important policy variables in the regression that can be used to affect cognitive development. To allow these variables to have different effects on children from different ethnic backgrounds, interaction variables (BLKIND, BLCKHS, and MEXHS) have been included. To determine the effect of Head Start on a Black child, the coefficients for Head Start and the Black-Head Start interaction term must be added. The coefficient of kindergarten for whites and Mexican-Americans (which was constrained to be the same because there were few Mexican-Americans) is positive in both the full-year and summer samples, but there is a great difference in the magnitude of the coefficients. Judging from other results, we suspect that the true effect lies somewhere between the two values. Our data indicate that kindergarten has almost a zero effect for Black children; no good explanation for this finding has been developed. It is possible that the selection process for kindergarten works differently for Blacks than it does for whites. The coefficients for kindergarten are subject to the same potential bias problem as Head Start, only more so. This is because no effort was made to match the kindergarten and nonkindergarten children, and discriminant analysis indicated that the kindergarten children come from significantly more advantaged backgrounds for all of our samples.

Head Start has a negative, insignificant coefficient for white children for both the full-year and summer samples. The Head Start-ethnicity interaction variables are positive in both samples, and the Black-Head Start interaction is significant at the 5 percent level for the summer sample. The net effect of Head Start for Blacks is 1.6 ITPA points for the summer sample and .82 ITPA points for the full-year sample. It is surprising to find a smaller effect for full-year than summer programs, and we suspect that this difference may be spurious. For Mexican-American children, summer Head Start has a positive effect (but smaller than for Black children), and full-year Head Start has a negative, insignificant effect.

To interpret the practical significance of the regression coefficients it is necessary to understand what a change of one point on the cognitive test means for the child. In the ITPA manual Kirk et al. (1968) offer two interpretations for the ITPA score. For children 5 to 7 years old, the manual suggests that an increase of one point on the ITPA score is approximately equal to a gain of three months of psycholinguistic age or three months of mental age. For children 5 to 7 years old a gain of one point on the ITPA is therefore equal to a gain of four to five IQ points. Thus, the net effect of summer Head Start for Blacks can be expressed as either 1.6 ITPA points, 4.8 months of mental age growth, or a 6.4 to 8.0 gain in IQ points. As IQ is the most widely used measure of mental ability, specifying the gains in an IQ metric allows for the easiest interpretations of

Head Start effects. Regressions using the Metropolitan Readiness Test (MRT) as the dependent variable are not presented here because the coefficients have the same sign as when the ITPA is used, and because the MRT offers no way to easily interpret the value of the gains.

In Table 5 a summary of the effects of Head Start and kindergarten when the ITPA is used as the dependent variable is presented for all ten samples. The Westinghouse researchers envisioned that the second and third grade samples could be used to test for the decay of the Head Start effect or the possible presence of a "sleeper" effect where the gains are not manifested until one or more years after the Head Start experience. However, it is dangerous to interpret the results in this manner because the study was not longitudinal, but instead was a series of cross-section samples. First, the selection procedures used by the Head Start centers may have changed over the three-year program; this might lead to noncomparable samples. More importantly, the Head Start programs may have changed over the period--hopefully improved--and a larger effect for the first grade children may be indicative of changes in the programs rather than decay of the effects.

For three of the four first grade samples analyzed, Head Start has a positive effect for Black children, and these effects are equivalent to four to ten IQ points; only for the summer, mother only sample is the effect negative for Blacks (and then it is insignificant and almost zero). Thus, the immediate impact of Head Start for Black children is quite favorable. For white children in the first grade Head Start is effective only when the father is absent from the home. For the second and third grade samples the effect of Head Start is not great for Black children

TABLE 5

Summary of Effects of Head Start and Kindergarten for All Regressions with ITPAMN as the Dependent Variable

Grade	Type of Program	Parents Present	Observations	Head Start for Whites	Head Start for Blacks	Kindergarten for Whites	Kindergarten for Blacks
1	full year	both	206	-1.22	.82	3.47**	-.09**
1	full year	mother	67	3.86*	1.33	5.21*	2.09
1	summer	both	623	-.36	1.63**	.86**	.27
1	summer	mother	143	2.62**	-.82	1.22	.42
2	full year	both	218	-.52	.08	3.57**	.77**
2	full year	mother	75	1.18	1.70	1.98	1.29
2	summer	both	635	-.48	-.57	1.46**	-.10**
2	summer	mother	134	1.11	.51	1.94**	1.60
3	summer	both	426	-.02	.58	1.97**	1.70
3	summer	mother	134	1.52	-.48	1.23	.52

*Statistically significant at the 10 percent level for white coefficients.

**Statistically significant at the 5 percent level for white coefficients.

*Statistically significant at the 10 percent level for Black interaction coefficients.

**Statistically significant at the 5 percent level for Black interaction coefficients.

except for the grade 2 summer sample where only the mother is present. For white children Head Start exhibits the same pattern of effects as for the first grade--Head Start has a fairly high effect (greater than one point on the ITPA) when only the mother is present, and a very small, negative effect when both parents are present. If we discount heavily the findings for the second and third grade samples on the grounds that Head Start programs were not yet fully developed, one may conclude that Head Start is effective for Black children and for white children from mother-headed families. These results are consistent with the overall findings of the Westinghouse study, but by adding additional control variables and by stratifying the data to permit the effects of Head Start to vary across different types of children, we have discovered that Head Start may be appropriate only for certain groups.

Kindergarten consistently shows a strong, positive effect for white children in all samples, and there is no trend for Black children. However, because the children who attended kindergarten are from significantly more advantaged backgrounds, we would not advocate an expansion of kindergarten programs on the basis of the findings reported here.

Although the bias issue can not be settled for the Westinghouse data, several techniques have been employed to determine if the Head Start and control children differ significantly on socioeconomic and demographic characteristics that are correlated with cognitive development. Discriminant analysis was used to test the joint hypothesis that Head Start and control children differ significantly on the independent variables included in the regressions; only for the first grade, summer, both parents present sample is there a significant difference favoring the control group. An alternative procedure for measuring these differences is discussed in the Appendix. Support of the null hypothesis that the groups do not differ significantly on these variables adds credence to the belief that the regressions produce unbiased estimates of the effects of Head Start but are insufficient to guarantee unbiasedness. Besides the fact that certain selection procedures will produce unbiased estimates when the groups differ on their pretreatment cognitive development, we may not have the complete set of relevant variables available, e.g., parental attitudes toward education. Discriminant analysis was also employed to determine if there were significant background differences between children who have attended kindergarten and those who have not. For all ten samples it was discovered that children who attended kindergarten are from significantly more advantaged backgrounds as measured by such characteristics as income, family size, and parental education. Thus, it is likely that the coefficients for kindergarten are inflated.

Several modifications of the basic regression model were formulated to test some additional hypotheses of interest. It has been observed by some researchers (e.g., Herzog et al., 1972) that preschool programs may provide smaller benefits to the most disadvantaged children. Such findings may reflect differential learning abilities by SES, but they may also be an indication that the curricula of these programs are set at too high a level for the most disadvantaged children. To test the hypothesis that Head Start is equally effective for all children in the sample, a set of Head Start-SES interaction variables was added to the regression equations. The variables were formed by multiplying the Head Start dummy variable by each of the independent variables. The hypothesis that the entire set of

Head Start-SES interaction coefficients are equal to zero was then tested. The Head Start-ethnicity variables (BLCKHS and MEXHS) were not included in the test as they are part of the basic model. The null hypothesis was not rejected for any of the first grade samples. Because of these findings we have rejected the hypothesis that the effect of Head Start is dependent upon the socioeconomic status of the child's family.

A second extension of the model that was carried out for the first grade both parents present samples was to add neighborhood and Head Start center dummy variables to the regression equation. The neighborhood variables may add explanatory power to the model by controlling for differences in neighborhood environments. When the Head Start center variables are added to the model each center is considered as a separate treatment; we were thus able to test if the Head Start effects for the various centers are clustered about the overall effect or if there is a large variation. For these analyses only children from neighborhoods with at least four Head Start and four control children were included. The neighborhood variables were added first, and the set of coefficients is significant for both the summer and full-year samples when the ITPA mean is used as the dependent variable; the addition of these variables did not significantly change the coefficients of the previously included independent variables. The set of Head Start center variables was added to the regression equation with the neighborhood variables also included (so that these variables would not serve as a proxy for neighborhood effects) and then tested for significance. The set of coefficients for the center variables was not significant at the 10 percent level. Thus, there does not appear to be significant variation in the effects of the centers analyzed.

4. Conclusions and Policy Implications

Although the reanalyses presented above do not completely reverse the pessimistic findings of the Westinghouse study, they do indicate that Head Start may have a significant immediate impact on children from minority groups and white children from mother-headed families. The next logical step in this evaluation is to place a dollar value on the cognitive gains to determine if the benefits of Head Start exceed the costs. Although this decision must ultimately be made, it is beyond the author's ability to carry out such an imputation. To make such imputations one would prefer to have longitudinal data available so that the permanence of the effects can be judged. Unfortunately, most of the longitudinal studies conducted so far have indicated that whatever cognotive gains are accrued during a preschool program seldom last past the first year of formal schooling; this was the finding in the surveys of the literature carried out by Lois-Ellin Datta (1969) and Marian Stearns (1971) for the United States Department of Health, Education, and Welfare.

It is clear that not enough is known at the present time to declare that Head Start is either a "success" or "failure." The first issue that must be addressed is whether or not an immediate gain in IQ of 5 to 10 points is a sufficient output for a preschool program. If these benefits from Head Start are judged to be worth attaining, additional research must be carried out to determine why the effects dissipate so rapidly. The policy implications for sustaining initial gains will depend upon the reasons for the subsequent loss. If the initial gains are simply an artifact due to the change of environment or changes in the children's attitudes toward taking tests, then efforts should be made to develop programs which produce real gains or preschool programs should abandon the attempt to affect cognitive development and concentrate on other aspects of development. Programs with earlier and later intervention should also be tried in case the ages of three to five are not as "critical" as was once thought. If it is found that the gains fade because they are not reinforced in the public school system, efforts should be made to better coordinate preschool and primary school programs. A final possible explanation for the fading of cognitive gains is that the home environment of disadvantaged children does not suitably reinforce the gains of preschool programs such as Head Start; if this is the case then efforts can be made to alter the home environment or to place the child outside that environment for a longer period of time.

Any strong conclusion about Head Start or preschool educational programs in general would be premature at this time. We do not agree with the statement by Arthur Jensen (1969) that "Compensatory education has been tried and it apparently has failed." Our research has produced some evidence that Head Start may be most effective for those children that Jensen predicts can be helped least--Blacks, and white children from fatherless families. It must be remembered that preschool education for the disadvantaged is a relatively new field. Research should continue so that we may learn how to sustain and enhance the initial gains that are found. At the same time, we must realize the limitations of preschool education. As Jencks et al. (1972) have correctly claimed, education will not eliminate inequality in our society. It is possible, however, that Head Start and other preschool educational programs can play an important part in the education of disadvantaged children.

NOTES

[1]Although all benefits and costs should be considered in the evaluation of a social action program, the major objectives should be given the most consideration. Cain and Hollister (1969, p. 6) claim that "in general the measures of program outputs, which may be proxies for the ultimate tangible changes, such as income change, employment gain, and educational attainment." With regard to Head Start, John W. Evans (1969, p. 254) states:

> . . . while Head Start has objectives other than cognitive and effective change, these other objectives are in large part <u>instrumental</u> to the cognitive and affective objectives. That is, the program is attempting to improve children medically and nutritionally <u>in order to</u> make it possible to change them cognitively and motivationally.

[2]Cain and Watts (1970, p. 238) make the same point about determining the most efficient policy alternative. In their presentation, however, they suggest scaling the policy variables so that one unit of any policy variable has the same cost. If this procedure is used then the most efficient policy variable is the one with the largest regression coefficient.

[3]Goldberger (1972a) has shown that the true regression of Y on X is not linear when selection is made on the basis of X*. Because empirical work is generally run using linear approximations, using a linear regression reflects what will happen in actual experiments. In qualitative terms, the spurious treatment effect retains the same direction of bias, but since the within-group regressions are no longer parallel the treatment effect calculated in the nonlinear regression will be a function of X.

[4]The major problem with the affective instruments is that they were designed specifically for the Westinghouse evaluation and were not verified on large samples. The comments of Victor G. Cicirelli, the principal Westinghouse investigator, and several OEO officials cast doubt on the validity and robustness of the affective measures. Cicirelli, Evans, and Schiller (1970, p. 115) state that "our judgment about the affective findings should be tentative and this is the view the Westinghouse report took." In another article Evans (1970, p. 256) says "No great claims are made for the affective instruments."

[5]The Hollingshead Index is formed by coding educational attainment and occupational status on seven-point scales. The Index is formed by multiplying the occupational score by seven, multiplying the educational score by four, and taking the sum of the two products. Including the information in this manner constrains the coefficient for education to be four-sevenths of the coefficient for occupation.

There are, however, several weaknesses of discriminant analysis. If some differences favor the control group and others favor the Head Start children, there is no way to take these offsetting differences into account. Even if all of the evidence indicates that one group comes from a better background, discriminant analysis does not provide any information about what these differences will be in terms of cognitive development. To get a better idea of the differences in cognitive development that would have been observed if there had been no treatment, we have developed a two-stage procedure as an alternative to discriminant analysis. The first stage of the procedure is to regress the cognitive measure used as the posttest (such as the ITPA score) on all the indep-pendent variables except Head Start for the control group. This produces an estimate of the educational production function without Head Start. This fitted function is then used to form an imputed test score for observations in both groups; the imputed score is formed by the formula $YHAT = \sum_{i=1}^{k} \beta_i X_i$ for each of the observations. In the second stage of the procedure, the variable YHAT is regressed on the dummy variable for Head Start. The coefficient for the Head Start variable then relates the difference in the average imputed test scores for the two groups. For the first grade, summer, both parents sample when the ITPA score is used as the dependent variable in the first stage, the second stage regression is:

$$\underset{}{YHAT} = \underset{(183.51)}{19.300} - \underset{(-5.047)}{.764}HDSTRT.$$

There were 623 observations for the regression and the R^2 is .0394. Thus we predict that the Head Start children would have scored .764 ITPA points lower than the control children if they had not had Head Start, and this is roughly equal to 3.8 IQ points. This procedure does not tell us whether or not the

APPENDIX

Techniques for Measuring Pretreatment Differences

There are several approaches that can be used to determine if the Head Start and control groups differ significantly in their pretreatment socioeconomic status. Perhaps the simplest is to perform an analysis of variance for each trait of interest. The problem with such a technique is that there is no good way to combine the results to make an overall judgment. Another possibility is to use discriminant analysis to test the joint hypothesis that the two groups have the same means on the variables of interest. The discriminant analysis can be carried out by regressing the dummy variable for Head Start on the same set of independent variables included in the educational production function. In the limiting case where there is no relationship between the independent variables and group membership all the coefficients would be zero and the constant would be 0.5 (assuming equal sample size in the two groups). Because we are dealing with a 0-1 dependent variable, the fitted values of the dependent variable can be interpreted as the probability that a particular observation will be in Head Start rather than the control group; this is not precisely correct and more sophisticated techniques have been developed for making such estimates. The regression coefficients can then be interpreted as the change in the probability of membership in the Head Start group for a unit change in the independent variables. The F-statistic for the test that the entire set of coefficients (excluding the constant) are equal to zero is equivalent to the test of the hypothesis that the means of each variable are the same in the two groups.

previous estimates of the effect of Head Start are biased, however, because we do not know which selection procedure was used to assign the children to Head Start. The procedure may be helpful in determining how different the experimental and control groups are in ex post facto analyses in terms of the metric used in the posttest.

BIBLIOGRAPHY

Barnow, Burt S. "Conditions for the Presence or Absence of a Bias in Treatment Effect: Some Statistical Models for Head Start Evaluation." Madison: Institute for Research on Poverty, 1972. Discussion Paper 122-72.

Barnow, Burt S. "The Effects of Head Start and Socioeconomic Status on Cognitive Development of Disadvantaged Children." Ph.D. dissertation, University of Wisconsin, 1973.

Bereiter, Carl. "Some Persisting Dilemmas in the Measurement of Change." In Problems in Measuring Change, edited by C. W. Harris, pp. 3-20. Madison: The University of Wisconsin Press, 1967.

Blalock, Hubert M., Jr. Causal Inferences in Nonexperimental Research. Chapel Hill: The University of North Carolina Press, 1961.

Bowles, Samuel. "Towards an Educational Production Function." In Education, Income, and Human Capital, edited by W. Lee Hansen. New York: National Bureau of Economic Research, 1970.

Cain, Glen G., and Hollister, Robinson G. "The Methodology of Evaluating Social Action Programs." Madison: Institute for Research on Poverty, 1969. Discussion Paper 42-69.

Cain, Glen G. and Watts, Harold W. "Problems in Making Inferences from the Coleman Report." American Sociological Review 35, No. 2 (April 1970): 228-242.

Campbell, Donald T. "Reforms as Experiments." American Psychologist 24, No. 4 (April 1969): 409-429.

Campbell, Donald T. and Erlebacher, Albert. "How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful." In Compensatory Education: A National Debate, Vol. 3 of Disadvantaged Child, edited by J. Hellmuth, pp. 185-210. New York: Brunner/Mazel, 1970.

Campbell, Donald T. and Erlebacher, Albert. "Reply to the Replies." In Compensatory Education: A National Debate, Vol. 3 of Disadvantaged Child, edited by J. Hellmuth, pp. 221-225. New York: Brunner/Mazel, 1970.

Campbell, Donald T. and Stanley, Julian C. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally & Co., 1963.

Cicirelli, Victor G. "Head Start: Brief of the Study." In Britannica Review of American Education. Vol. I., edited by David G. Hays. Chicago: Encyclopedia Britannica, 1969.

Cicirelli, Victor G. "The Relevance of the Regression Artifact Problem to the Westinghouse-Ohio Evaluation of Head Start: A Reply to Campbell and Erlebacher." In Compensatory Education: A National Debate, Vol. 3 of Disadvantaged Child, edited by J. Hellmuth, pp. 211-215. New York: Brunner/Mazel, 1970.

Cicirelli, Victor G.; Evans, John W.; and Schiller, Jeffry S. "The Impact of Head Start: A Reply to the Report Analysis." Harvard Educational Review 40, No. 1 (Winter 1970): 105-129.

Cicirelli, Victor G. et al. The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development. Vol. I and Vol. II. A report presented to the Office of Economic Opportunity pursuant to contract B89-4536, June 1969. Westinghouse Learning Corporation, Ohio University.

Cicirelli, Victor G. et al. "Performances of Disadvantaged Primary-Grade Children on the Revised Illinois Test of Psycholinguistic Abilities." Psychology in the Schools 8, No. 3 (1971): 240-246.

Coleman, James S. et al. Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office, 1966.

Cramer, J. S. "Efficient Grouping, Regression and Correlation in Engel Curve Analysis." Journal of the American Statistical Association 59, No. 305 (March 1964): 233-250.

Datta, Lois-Ellin. A Report on Evaluation Studies of Project Head Start. Washington, D.C.: Project Head Start, Office of Child Development, U.S. Department of Health, Education, and Welfare, 1969.

Evans, John W. "Head Start: Comments on the Criticisms." In Britannica Review of American Education. Vol. I., edited by David G. Hays. Chicago: Encyclopedia Britannica, 1969.

Evans, John W. and Schiller, Jeffry. "How Preoccupation with Possible Regression Artifacts Can Lead to a Faulty Strategy for the Evaluation of Social Action Programs: A Reply to Campbell and Erlebacher." In Compensatory Education: A National Debate, Vol. 3 of Disadvantaged Child, edited by J. Hellmuth, pp. 216-220. New York: Brunner/Mazel, 1970.

Goldberger, Arthur S. "Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations." Madison: Institute for Research on Poverty, 1972. Discussion Paper 123-72. (1972a)

Goldberger, Arthur S. "Selection Bias in Evaluating Treatment Effects: The Case of Interaction." Madison: Institute for Research on Poverty, 1972. Discussion Paper 129-72. (1972b)

Grotberg, Edith S. "Review of Research 1965-1969." Washington, D.C.: Office of Economic Opportunity, 1969.

Herzog, E. et al. "But Some Are Poorer Than Others: SES Differences in a Preschool Program." American Journal of Orthopsychiatry, 42. No. 1 (January 1972): 4-22.

Hilderth, Gertrude H.; Griffiths, Nellie L.; and McGauvran, Mary E. Manual of Directions: Metropolitan Readiness Tests. U.S.: Harcourt, Brace & World, Inc., 1966.

Hollingshead, August B. Social Class and Mental Illness. New York: John Wiley & Sons, Inc., 1958.

Jensen, Arthur R. "How Much Can We Boost IQ and Scholastic Achievement?" Harvard Educational Review 39, No. 1 (Winter 1969): 1-123.

Kmenta, Jan. Elements of Econometrics. New York: The Macmillan Company, 1971.

Light, Richard J. and Smith, Paul V. "Choosing a Future: Strategies for Designing and Evaluating New Programs." Harvard Educational Review 40, No. 1 (Winter 1970): 1-28.

Levitan, Sar A. The Great Society's Poor Laws: A New Approach to Poverty. Baltimore: The Johns Hopkins University Press, 1969.

Lord, Fredric M. "Large-Sample Covariance Analysis When the Control Variable is Fallible." Journal of the American Statistical Association 55 (1960): 307-321.

Lord Fredric M. and Novick, Melvin R. Statistical Theory of Mental Tests. Reading, MA.: Addison-Wesley, 1968.

Madow, William G. "Project Head Start, a National Evaluation: A Methodological Critique." In Britannica Review of American Education, Vol. 1, edited by David G. Hays. Chicago: Encyclopedia Britannica, 1969.

Stearns, Marian S. Report on Preschool Programs: The Effects of Preschool Programs on Disadvantaged Children and Their Families. Washington, D.C.: Office of Child Development, U.S. Department of Health, Education, and Welfare, 1971.

Watts, Harold W. and Horner, David L. "The Educational Benefits of Head Start: A Quantitative Analysis." Madison: Institute for Research on Poverty, 1968. Discussion Paper 14-68.

White, Sheldon H. "The National Impact Study of Head Start." In Compensatory Education: A National Debate, Vol. 3 of Disadvantaged Child, edited by J. Hellmuth, pp. 163-184. New York: Brunner/Mazel, 1970.

Wholey, Joseph S. et al. Federal Evaluation Policy: Analyzing the Effects of Public Programs. Washington, D.C.: The Urban Institute, 1971.

Williams, Walter and Evans, John W. "The Politics of Evaluation: The Case of Head Start." The Annals of the American Academy 385 (September 1969): 118-132.