14-68

THE EDUCATIONAL BENEFITS OF HEAD START: A QUANTITATIVE ANALYSIS

Harold W. Watts
and
David L. Horner

The authors are respectively Professor of Economics and Director of the Institute for Research on Poverty, and Research Assistant in the Institute for Research on Poverty.

This research was supported by funds granted to the Institute for Research on Poverty pursuant to the provisions of the Economic Opportunity Act of 1964. We wish to thank the Office of Economic Opportunity, the Planning Research Corporation, and the Census Bureau for providing the data necessary for this study. John Conlisk, Robinson Hollister and Arthur Goldberger offered valuable suggestions and read the drafts at several stages.

Preface

This study is designed to add to the decision-maker's knowledge about Head Start. It does not consider the whole range of Head Start's benefits and costs, but is limited to an examination of its educational component. The study seeks to determine the mean effect on a child's score on an educational achievement test and a social readiness test produced by increments of time spent in Head Start, while simultaneously controlling for other factors, especially age, that enter into the determination of the score. The results, while less powerful than one would desire for strong assertions, do indicate that lengthening the child's participation in Head Start increases his educational achievement level. More definite confirmation was found from the proposition that extra weeks in Head Start will enhance social readiness.

This revised edition of the earlier discussion paper bearing the same name includes a section on the cost and benefits of Head Start as well as an extention of the previous analysis. The estimates for the Peabody Picture Vocabulary Test indicate that the increase in the educational achievement level due to Head Start is about half of that due to age effects alone. Given that the mean educational achievement level of Head Start children is about one year behind that of the average children of comparable age a program lasting two years would be needed to close the gap, at a cost of around $1600 for the "education related" components alone. On the simple assumptions used for this calculation, a shorter and cheaper program would fill a correspondingly smaller part of the gap.

I PURPOSE OF THE INVESTIGATION

Head Start is a multi-objective program designed for underprivileged preschool children. Central to the several objectives such as detecting and correcting social, medical, dental and nutritional problems, there is the educational objective. One of the reasons that children fall behind and eventually drop out of school is their inability to keep pace with the general class level. If, as is thought, it is possible to identify the children most likely to fall behind, perhaps something can be done to remove their handicaps. Head Start endeavors to prepare the child for the primary school environment by attempting to offset or compensate for the deficient or actually harmful conditions to which the child has already been exposed. This paper aims at an estimate of Head Start's effectiveness in improving the child's educational and social readiness to interact or compete with the "average" child.

To obtain such an estimate it is necessary to have, first, some measure of the child's educational and social readiness, and second, a variable which quantifies the child's "exposure" to Head Start. Assuming we have a test on which the child's score improves as his readiness for school increases and a variable such as number of weeks the child has been in Head Start, then the problem can be stated: By how much does the child's test score improve as the number of weeks in Head Start increases, given control of age and other characteristics which might affect the score?

A secondary measure of the impact of Head Start can be obtained from estimates of the amount that the child's test score improves as the total time per day in a Head Start class increases.

II TESTS AND VARIABLES

The two tests that were used to determine the child's educational and social readiness were the Peabody Picture Vocabulary Test (PPVT) and the Vineland Social Maturity Scale (VSMS). In the PPVT a word is spoken to the child who then chooses from among four pictures the one that corresponds to the word. Insofar as the tester is consistent in administering the test the score is objective. The VSMS is not a test given to the child but rather a questionnaire in which a person knowing the child (usually a teacher) rates his ability on the basis of a number of social criteria. This test suffers not only from measurement errors but also from the differences in the frame of reference of evaluators in different parts of the country. In order to correct for the latter problem it would be necessary to adjust the mean level of scores at each location.

Table I contains the list of dependent and independent variables and their sample means, variances and coefficients of variation. Table II displays the simple correlations of each regressor with the two dependent variables and the number of weeks and hours. These statistics prove useful in spotting biases occurring in regressions containing only subsets of the original regressors.

III DATA AND SAMPLE

The sample of Head Start children taken from Head Start grants for the 1965-66 academic year was drawn by the Planning Research Corporation

Table I

VARIABLES AND THEIR SAMPLE STATISTICS

	Dependent Variable	Abbreviation	Sample Means	Sample Std. Dev.	Coefficients of Variation
A.	Peabody Picture Vocabulary Test Raw Score	PPVT	40.83	11.21	0.27
B.	Vineland Social Maturity Test Raw Score	VSMS	56.36	6.31	0.11
	Independent Variables (a)				
1.	Number of weeks child was in Head Start Program at time of testing.	Weeks	19.19	7.45	0.39
2.	Hours per day child was in class.	Hours	3.81	1.34	0.35
3.	Previous PPVT Test given to child.	Prev. PPVT	0.08	0.28	3.31
4.	Previous pre-school training.	Prev. Train.	0.35	0.48	1.37
5.	Age in months.	Age	60.81	9.14	0.15
6.	Race (Standard is caucasian)				
	6a. (Negro)	R_n	0.53	0.50	0.95
	6b. (Other than Negro or Caucasian)	R_o	0.01	0.11	8.86
7.	Sex (female)	Sex	0.48	0.50	1.03
8.	Father absent from home.	Father	0.28	0.45	1.62
9.	Mother absent from home.	Mother	0.02	0.13	7.53
10.	Number of persons in the household.	Family size	6.71	2.52	0.38
11.	Number of Siblings between 0-15 years of age.	Siblings	3.41	2.17	0.63
12.	Mother's education level (7 point scale).	$Ed._m$	5.06	1.22	0.24
13.	Family Income (9 point scale).	Income	4.14	1.87	0.45
14.	Family receives welfare	Welfare	0.29	4.53	1.57
15.	Number of persons per sleeping roon	Per Room	2.45	1.04	0.42
16.	Number of Consumer Durables in Home (Unweighted sum of car, radio, television, and telephone).	Durables	3.15	9.23	0.29

Table I (continued)

VARIABLES AND THEIR SAMPLE STATISTICS

Independent Variables (a)	Abbreviation	Sample Means	Sample Std. Dev.	Coefficients of Variation
17. No Newspaper regularly present in home.	Newspaper	0.20	0.40	1.98
18. No running water in the home.	Water	0.06	0.23	4.08
19. Degree of Urbanization 4 point scale (Suburban location is the top of the scale, farm location, the bottom.	Urban	3.00	0.51	0.17
20. Regional Variables (Standard is the North-east region.				
20a. North Central Region	N. C.	0.23	0.42	1.85
20b. Southern Region	S.	0.19	0.40	2.04
20c. Western Region	W.	0.29	0.46	1.55

(a) Variables 3,4,6,7,8,9,14,17,18, and 20 are dummy variables in which the alternative would, in combination with the variable, represent a mutually exhaustive set. Thus, the coefficient of Sex (female), used as an independent variable in any regression, would represent the expected difference in score resulting from a child's being a female rather than a male. Likewise, variable 6a is a dummy measure of a child's being a Negro rather than a Caucasian and 6b is a dummy measure of his being other than Negro or Caucasian rather than a Caucasian. Thus, if one wanted to find the expected difference in a particular dependent variable, say PPVT Test Score, between Negroes and other non-Caucasians, one would calculate the difference in the coefficients of the dummy variable as they were determined in a regression which included both.

Table II

VARIABLES AND THEIR SAMPLE CORRELATIONS WITH EXPOSURE AND TEST MEASURES

Variables	Simple Correlation With PPVT	Simple Correlation With VSMS	Correlation With Weeks	Correlation With Hours
PPVT	1.000	0.406	-0.048	-0.080
VSMS	0.406	1.000	0.032	-0.079
Weeks	-0.048	-0.032	1.000	0.211
Hours	-0.080	-0.079	0.211	1.000
Prev. PPVT	-0.125	-0.118	-0.199	-0.132
Prev. Train.	-0.085	-0.014	-0.059	0.019
Age	0.469	0.395	-0.181	-0.151
R_n	-0.392	-0.155	0.173	0.029
R_o	0.057	0.060	-0.060	0.026
Sex	-0.053	0.087	0.006	-0.024
Father	-0.060	-0.014	-0.006	-0.037
Mother	-0.082	-0.100	-0.039	-0.009
Family size	-0.017	-0.014	-0.070	-0.065
Siblings	-0.011	-0.003	-0.007	0.037
$Ed._m$	0.040	-0.039	0.121	0.033
Income	0.072	0.050	0.070	-0.055
Welfare	-0.102	-0.064	-0.077	-0.061
Per Room	-0.073	-0.040	-0.131	-0.031
Durables	0.138	0.075	-0.009	-0.064
Newspaper	-0.059	0.042	-0.109	-0.064
Water	-0.015	-0.073	-0.063	-0.027
Urban	-0.024	-0.082	-0.022	-0.015
N.C.	0.031	-0.018	0.167	-0.272
S.	0.077	0.205	-0.199	-0.073
W.	0.081	0.060	-0.376	-0.224

of Los Angeles for use in its own analysis.[1] Data including test scores, weeks, hours, age, sex, and race were collected for 963 children by trained personnel. In addition, the parent or guardian of each child received a family information form from which all other variables were observed. The collection of this family information was the responsibility of the Census Bureau, which mailed out the forms. A follow-up letter was sent to families who did not return the questionnaire within a certain deadline. Family information was collected for 655 children. For the purposes of this study the sample was further reduced to 635 observations by eliminating those with missing information.

Because of the income eligibility criterion and the voluntary nature of Head Start, a sample drawn solely from Head Start children will not be representative of all preschool children. One indication of this is the sample PPVT mean of the data is 40.83 with a mean age of 61 months whereas the population mean of all preschool children of 61 months is estimated to be 50. The children in the sample were on the average lacking about 14 months in their educational development. On the other hand the sample VSMS mean of 56.5 is the same as the estimated population mean of all preschool children of 61 months. This is probably due to the more relative nature of scoring on the VSMS. Thus although the sample is not representative of preschool children in general it does represent a significant portion of the poor population whose families are receptive to social programs designed to supplement a deficient environment.

---

[1]More detailed information concerning how the sample was drawn and how the data were collected can be found in William D. Commins, H. Russel Cort, Naomi H. Henderson, and Ruth Ann O'Keefe, " A Study of the 1966 Full-Year Head Start Programs," PRC D-1268, prepared for the Office of Economic Opportunity, 19 September 1966.

Certain of the variables are also subject to measurement errors and biases because the family information form was completed by untrained people who were reporting on themselves. No attempt has been made to correct for such errors. Further discussion of the problems connected with some variables appears below.

IV THE BASIC REGRESSIONS

Test of the Hypothesis

The basic regressions included are the independent variables listed in Table I. Table III displays the regression coefficients and their standard errors. In regression A the dependent variable is PPVT Raw Score. VSMS Raw Score is the dependent variable in regression B.

In a one-tailed *t* test, the coefficient for Weeks in regression A is not significantly different from zero at the 5 percent level. The *t* ratio, 1.14 is just compatible with significance at the 15 percent level.[2] The coefficient for Weeks in regression B is significantly different from zero at the 5 percent level but not at the 1 percent level. According to these tests the evidence is consistent, at usual significance levels, with no change in PPVT scores attributable to Head Start exposure. The results are *not*, however, consistent with a similar absence of effect for the VSMS score.

The coefficient for the number of hours was negative but of trivial magnitude in regression A and positive but not significant at the 5 percent level in regression B. However, two problems are related to the interpretation of this variable. First, most of the man-hours-per-day

[2] In other words, if there were truly no effect of Head Start, an estimate as large as this one might be produced one time out of six simply by chance.

Table III

REGRESSION COEFFICIENTS AND THEIR STANDARD ERRORS[a]

Independent Variable	Regression A (PPVT Raw Score)		Regression B (VSMS Raw Score)	
	Coefficient[b]	Std. Error	Coefficient[b]	Std. Error
Constant Term	2.704	5.381	37.489	3.279**
1. Weeks	0.066	0.059	0.069	0.036*
2. Hours	-0.071	0.333	0.217	0.203
3. Prev. PPVT	-0.993	1.464	-1.5734	0.892
4. Prev. Train.	-0.626	0.787	0.229	0.480
5. Age	0.565	0.051**	0.254	0.031**
6a. R_n	-5.854	0.889**	0.318	0.542
6b. R_o	-1.928	3.367	-0.115	2.052
7. Sex	-1.463	0.740*	0.951	0.451*
8. Father	1.837	1.161	1.741	0.708*
9. Mother	-4.516	2.932	-3.846	1.787*
10. Family size	0.151	0.315	0.010	0.192
11. Siblings	-0.303	0.353	-0.033	0.215
12. $Ed._m$	1.427	0.344**	0.148	0.209
13. Income	0.027	0.251	0.219	0.153
14. Welfare	-1.068	1.065	-0.701	0.649
15. Per Room	-0.952	0.432*	-0.354	0.263
16. Durables	0.713	0.482	0.398	0.294
17. Newspaper	-1.320	0.980	0.546	0.597
18. Water	-1.056	1.679	-3.682	1.023**
19. Urban	-0.151	0.727	-1.085	0.443*
20a. N. C.	0.621	1.248	1.558	0.760*
20b. S.	-1.856	1.457	3.325	0.888**
20c. W.	1.778	1.269	2.973	0.773**

(a) The R^2's for the two regressions were 0.355 and 0.245 respectively.

(b) One and two asterisks denote significance at the 5% level and 1% level respectively. Except for variables 1, 2, and 5 values of t for two-tailed tests of significance are applied.

programs were in Philadelphia, which formed about 20 percent of the sample, and the variable Hours was in part acting as a proxy for locational differences. This will be discussed later with other problems connected with the location variables.

Second, there is a possibility that the variable representing hours of class time per day was measured incorrectly for a part of the sample. The actual data were recorded in terms of hours of center operation per day. In addition to the possibility that "hours of center operation" was longer than "class hours," there is a strong possibility that the centers with the longest hours of operation held two classes per day with two different sets of children. The number of center program hours ranged from 2.5 to 7.0. It might be possible to verify the accuracy of the measure by investigating the grant documents. In lieu of this, additional regressions were run excluding all observations in which the class hours were greater than four to eliminate the probability that centers held two classes per day. These regressions, containing 476 observations, yielded coefficients and standard errors for hours equal to .215 and .948 respectively for regression A and -.803 and 5.969 for regression B. Hence this device for eliminating measurement error reduces variation in hours so much that all precision is lost in estimating the coefficients.

<u>Other Variables</u>

The control variables i.e., all except for Weeks and Hours, can roughly be classified into four groups: (1) attributes of the child, variables 3-7; (2) family attributes, variables 8-12; (3) family income and consumption patterns, variables 13-18; and (4) location, variables 19-20. The variables in the second and third groups were derived from the family information forms, as were variables 4 and 19, and were less subject to control by trained personnel. With respect to variables 6b and 9--both dummy variables--there was little variation, only eight pupils were neither Negro nor Caucasian, and only thirteen were living apart from their mothers.

The group 1 variables are self explanatory and only a few comments are necessary. Care should be taken in interpreting the coefficients of the control variables because of the selective nature of the sample. For example, the effect of being Negro as measured by the coefficient of R_n is significant in regression A but not in regression B. This may reflect the fact that the PPVT test measures absolute scores based on some universal standard whereas the VSMS measures relative differences perceived by the test administrator. Thus a child who is at the head of his rural Mississippi class might be expected to score lower on the PPVT test than on the VSMS test. This interpretation points up the need to control more closely for varying locational environments. One might say that the use of the R_n variable has already accounted for part of this problem, since it is a proxy for environment; but this is only partly true since R_n is a proxy for other things such as bias in PPVT and VSMS test design (with respect to this note the significance of sex in both regressions A and B). Also there is not a one-to-one correspondence between Negroes and poor locational environment nor between Caucasians and good locational environment. But, whatever the reasons, the estimates suggest that Negro children are about a year behind white children in PPVT levels.

Previous exposure to the PPVT test or to pre-school programs did not appear to make a significant difference in a two-tailed test at the 5 percent level. With respect to the former (where both coefficients were negative by more than one standard deviation) it should be pointed out that most of the fifty-three children having previously taken the PPVT test were clustered together.

Of the group 2 variables the coefficients of family size and siblings were not significant. Probably the most surprising fact in both regressions was that the children from fatherless homes scored higher on both the PPVT test and VSMS test. In the case of the latter the difference was significant at the 1 percent level. The mother's educational level was a significant variable in regression A but not in regression B.

Group 3 variables produced coefficients with plausible signs. The children from families on welfare, without running water, with a large number of persons per sleeping room, with fewer durables, or with lower incomes had lower average test scores.

None of the group 3 variables was significant by itself with the exception of Per Room for regression A and Water for regression B. However, the measurement of the Per Room variables was evidently subject to great error. In a number of cases respondents noted more rooms used for sleeping than there were people in the household; in some cases the total number of sleeping rooms was high enough to suspect that they had misread the question and had included all the rooms in the house.

An initially puzzling factor was that the coefficient of Income was not significant in either regression. The nine categories of family income and the percentage distribution of the observations were as follows:

Family Income	Percentage of Observations
1. Under $1,000	7.9
2. $1,000-$1,999	14.7
3. $2,000-$2,999	14.0
4. $3,000-$3,999	20.6
5. $4,000-$4,999	17.3
6. $5,000-$5,999	14.0
7. $6,000-$7,999	8.7
8. $8,000-$9,999	1.4
9. $10,000 and over	1.4

With this distribution one would have expected differences in scores due to varying income. Several factors are at work here. One is that in other simpler regressions income acts as a proxy for many of the variables that we have been able to control. Thus, like race, income is highly correlated with environmental factors. If enough of these other factors are included then income loses its significance.

One might still expect income to be significant in its own right. It might not be significant, however, if observations in which relatively high regional incomes were associated with low income regions caused by a uniform poverty standard predominate. For example, a child from a family with an income of $4000 in a small town in Mississippi might be expected to score higher than a child from a family with an income of $4000 in New York because of differences in relative economic status. Still another reason for the nonsignificance of income might lie in the voluntary nature of the sample. Thus families with relatively high incomes might choose to enroll their children in Head Start only if they were convinced that the children had unusual needs for such a program. Those with low income would enroll their children only if the parents were strongly motivated to secure a good education for their children.

Locational variables were poorly specified in this study. One problem previously mentioned in connection with the variable Hours and the Philadelphia program is that no variable was used to control specifically for overrepresentation in several geographic locations. The regional variables correspond only approximately to areas of homogeneous levels of living. A better specification might be to use individual dummy variables for areas in which a significant number of observations were drawn, such as Chicago, Los Angeles, and Philadelphia.

The Urban variable was only crudely controlled. The categories were that the family lived (1) on a farm, (2) in the country, but not on a farm, (3) in a city or town, or (4) in a suburb or on the outskirts of a city or town. Although one might expect this to be the "proper" ordering, for purposes of control it might have been preferable to use individual dummy variables for each category. But there was also a problem of measurement error connected with this variable. The untrained family member filling out the form might easily have confused category one and two, two and four, three and four, or in fact any two categories. Treatment by transforming it into a series of dummies might have yielded only spurious accuracy.

V SUPPLEMENTARY REGRESSIONS

Several additional sets of regressions were fitted to subsets of the variables in the main regression. In all cases the same dependent variables (the PPVT and VSMS test scores) were used. Results of these regressions are listed in Table IV and the Appendix. To the basic regressions, Weeks and Hours, variables were added by groups in order to determine their collective significance. The F-statistics for the separate addition of groups 1 through 4 to the basic independent variables are listed in columns one and two of Table IV (the 1 percent significance levels are listed in parentheses). Columns three and four present the F-statistics for separately adding groups 2 through 4 to the combination of basic independent variables plus the group 1 variables. Group 1 variables had by far the greatest collective significance, and group 2 variables were not collectively significant in any of the regressions. Group 3 variables were alternately significant and not significant

Table IV

F-LEVEL FOR THE ADDITION OF BASIC GROUPS OF VARIABLES

	Addition to Basic Independent Variables		Addition to Basic Independent Variables plus Group 1	
	PPVT	VSMS	PPVT	VSMS
Group 1 (Attributes of Child)	39.32 (2.85)	20.30 (2.85)	-- (--)	-- (--)
Group II (Family Attributes)	1.68 (3.06)	1.75 (3.06)	0.91 (3.06)	2.30 (3.06)
Group III (Income-Consumption Patterns)	3.24 (2.85)	1.85 (2.85)	0.62 (2.85)	3.95 (2.85)
Group IV (Location)	4.42 (3.36)	14.02 (3.36)	0.54 (3.36)	5.10 (3.36)

Table V

SUPPLEMENTARY REGRESSIONS

	PPVT (Basic + Group 1)		VSMS (basic + Groups 1, 4)	
	Coef.	Std. Error	Coef.	Std. Error
Constant	16.157	3.689	38.705	2.789
Weeks	0.099	0.055	0.083	0.036
Hours	-0.286	0.296	0.172	0.206
Prev. PPVT	-0.843	1.466	-1.692	0.887
Prev. Train.	-1.425	0.808	0.238	0.483
Age	0.458	0.048	0.253	0.032
R_n	-5.145	0.873	0.628	0.532
R_o	-0.498	3.473	0.724	2.069
Sex	-1.410	0.767	0.848	0.457
Urban			-0.809	0.448
N.C.			1.659	0.761
S			2.393	0.860
W			3.145	0.766

depending on whether group 1 variables were already in the equation. Group 4 variables were significant in both cases for VSMS scores but the case is not so clear cut with respect to PPVT scores.

Table V presents the results when sets of variables are chosen for entry into the regressions according to their statistical significance. For PPVT scores the basic variables plus the group 1 variables are the only ones which survive. In this regression the Weeks coefficient is larger and more precisely estimated than before. However, the coefficient for hours is negative by one standard deviation. For the VSMS scores the basic variables plus group 1 and group 4 pass the test for inclusion. In this equation the t-ratio for the variable weeks is 2.28. An additional factor lending credibility to these results is that with the exception of previous preschool training the included regressors correspond to those variables which were collected by trained personnel.

Absence of any substantial systematic relation of the "background" variables in groups 2 and 3 (Family Attributes and Income-Consumption Patterns) to either measure of performance requires some comment. Intentional selection of children from deprived backgrounds has reduced the range of variation in these factors. A sample drawn mainly from the low socio-economic strata should not be expected to display a strong relation between family background factors and intellectual or motivational levels of the sort found in the Coleman report. Hence the findings of this study should not be regarded as inconsistent with the well established relation of background and achievement over the entire population of children.

VI COSTS AND BENEFITS

Several problems arise when attempting to make cost comparisons using the present body of data. The gross cost data for the 1966 part-day program is not broken down by length of program. The benefit analysis contained herein is restricted by the assumption of a constant benefit per unit of time. Thus the main result which can be inferred is that of a constant benefit per dollar accruing with a lengthening of a Head Start Program. This information can be used to determine the optimal time period for Head Start if we know the alternative benefit cost ratios. If the Head Start benefit cost ratio is higher than that of all alternatives, then Head Start is to be preferred until some time when an alternative program demonstrates a superior ratio. For example, the logical time to end a Head Start may be when the child enters first grade provided the benefit per unit cost of the usual formal schooling surpasses that of Head Start. This is not to say that other programs will not be needed for the underprivileged child in conjunction with his formal schooling but this is another question to be studied in relation to the retention and reinforcement of Head Start benefits.

Costs

The available data on Head Start gives different cost figures for summer part-day programs, full-year part-day programs and full-year full-day programs. Table VI displays a summary of the average costs per child per month both for the total program and for those components directly related to the educational objective.

One might suppose that the per child costs of the two-month summer program could be used with the per child costs of the full-year (10 months)

Table VI

Program	Total Cost per Child per Month	Cost for Daily Activities and Administration per Child per Month
1. 1965 Summer Part-Day Program, 2 months	75	61.80
2. 1966 Full-Yr., Full Day, 12 months	105	77.20
3. 1966 Full-Yr., Part-Day Program, 10 months	90	65.00
4. 1966 Summer Part-Day Program, 2 months	96	56.60
5. 1967 Full-Yr., Full Day, 12 months	115	83.60
6. 1967 Full-Yr., Part-Day Program, 10 months	97	68.80
7. 1967 Summer Part-Day Program, 2 months	100	58.90

program to obtain a crude marginal cost curve, but this approach has several drawbacks. The cost data do not clearly indicate whether the short-session costs are greater or less than the long-session costs.

In addition, these costs are not entirely comparable over time because of changing standards and because the demand relative to the supply of educators is probably less during the summer months (although government expenditure policy may not take this into account). The one relationship that does seem to hold is that the cost for the educational component is less in the summer than the full-year program. But in any case the benefit estimated here is based on the full-year program and the benefit per month for the summer program may well be smaller.

Further analysis of the cost portion of the benefit cost ratio of Head Start might be obtained by introducing program cost as a variable in a multiple regression analysis such as the one used in this paper. For the longer run it would be possible to design experiments which would include cost as a variable.

Benefits

There are a number of preliminary remarks that are in order prior to an attempt to interpret the results of the regression analysis. First with regard to the measures of performance used to guage the effect of Head Start, what is the relevance of each measure to subsequent success in school, and what is the relation between them. Two measures, available in the P.R.C. data, were used in the regression analysis, and no attempt has been made to combine these into a more comprehensive indicator. Such a combination would require knowledge or at least working hypotheses about the learning process well beyond the competence of the authors. It is hoped that others will be able to supply this crucial element.

In the sample used for this study, there was a substantial gap between mean PPVT scores and the national norm, and the regression indicated a quite imprecisely estimated effect of Head Start tending to close this gap. By contrast the VSMS ratings were no different from the norm and the regression estimated a beneficial effect with some reliability. If social maturity, as measured by the VSMS is "satisfactory" at or around the norm and further increments of negligible value relative to gains in the more cognitive skills measured by the PPVT, then most of the emphasis should be placed on the latter. But this is only one of many possible, and equally plausible, weightings from the point of view of the authors.

The present study has not attempted to explore many possible interaction effects of the Head Start program. The notion that Head Start, or more precisely weekly increments of Head Start, might be differentially effective on children of different ages (chronologically on in terms of development levels) has some *a priori* appeal and should be investigated. Differentials according to race, region, sex, etc. would be relevant in determining the distribution of benefits. A larger and more closely controlled sample would be necessary for an attempt to draw out such interaction. But because they have been neglected here, the results must be regarded as crude and tentative.

Finally, the results here relate, at best, to a proximate or immediate goal of the Head Start program. It is reasonable to ask what immediate effects can be detected in a child's performance which appear to be a result of the Head Start experience. But if after some interval, it is no longer possible to distinguish between a group of Head Start "graduates" and a comparable group who had no pre-school training, then those measured immediate benefits would be of little value in terms of ultimate objectives. While the improvement in test performance as used in this study may be accounted as a clear gain, we must remember that such gains can be lost again--perhaps through the same process that produced the initial discrepancy. There is, for this reason, a great need for evaluative studys which can follow the children from the early Head Start programs as they progress through the school system.

A careful evaluation of Head Start should compare the costs and benefits of alternative programs. Unfortunately hard evaluation evidence is more scarce for possible alternatives than it is for Head Start itself. If the analysis had yielded sharper estimates on the effects of a more intensive (longer hours) as compared to a more extensive (more weeks) program, there would have been some useful contrasts within the current range of Head Start variations. There should be more study of alternatives, those within the basic Head Start framework, those within the regular school system, and other possibilities such as programs for the very young, 10-hour day care, or a more complete replacement of the home as exemplified in the Kibbutz.

Although all of the above questions require more study, we can within limits derive some interpretation from the present results. In the case of the PPVT analysis, the basic regression shown in Table III estimates an increase of .066 raw score points as the <u>weekly</u> incremental effect of Head Start, and an increase of .565 as the <u>monthly</u> effect of age by itself. In both cases, these estimates are based upon cross-sectional differences, rather than on longitudinal observations. But if they are projected longitudinally they suggest that during a six-months period a child's score should rise by 3.39 points (6 x .565) simply because the child gets older, and an additional 1.72 points (.066 x 26) if the child has spent those six months enrolled in a Head Start program.[3] Thus, the point estimate indicates a 51% increase over the "normal rate of change" using this measure of competence.

Considering the standard error of the estimate of the weekly Head Start increment, which was equal to .059, one must acknowledge that the evidence leaves a substantial range of uncertainty about Head Start's effect. Multiplying .059 by 26 and adding and subtracting one standard deviation from the point estimate produces a range that stretches from no appreciable effect on the rate of development all the way to a doubling of that rate due to Head Start. In this sense, then, the findings do not provide an unequivocal basis for confidently projecting a beneficial effect.

Almost everyone would agree that if the actual effect is at the low end of the estimate we should not be spending money in this way. But the data are equally consistent with the true effect being at the high end of the interval--and such a doubling of the rate of development would surely be regarded as worth quite a bit of money.

There is evidence from external data on the normal development of children as measured by the PPVT test. In the sample analyzed here, the mean score was 41 points, and the average age of the children was 61 months. A normal score for children of that age is 50 points. In the normal pattern of development a score of 41 is achieved at the age of 47 months--showing these children to be about a year and two months behind the normal pattern.

---

[3]Using the more favorable estimate for the coefficient of weeks presented in Table V (.099 for weeks of Head Start vs. 0.458 for age) the indicated six month increase due to Head Start is 2.57 or 93% of the corresponding increase due to age (6 x .458 = 2.75). Note, however, that the combined effect (age + Head Start) is 5.32 as compared to 5.11 for the basic regression.

The normal rate of change over a six-month period varies with age--with fairly rapid changes for young children, diminishing as the child grows older. During the fourth year, the normal pattern suggests about 3½ or 4 points change each half year. During the fifth year the average change is 3 points per half year.

These results seem to agree well with the pure age effects estimated in our regression. The main regression results imply, moreover, that around two years of Head Start would be needed to eliminate the average deficiency in development noted for the sample of children we used for our analysis. If the more favorable regression results obtained in Table V are applicable then a year and a summer would eliminate the average deficiency in development.

The other score which was available from the PRC study was the Vineland Social Maturity Scale. The effect of Head Start on this measure was estimated with a slightly more satisfactory degree of precision. Translating the estimates into six-month changes once more, one sees that the point estimates for age and weeks of exposure suggest that Head Start more than doubles the rate of "maturation"--with a one-standard-deviation range going roughly from 50% increase to 150% increase. However, several facts concerning the normal social development of children as given in the VSMS manual are relevant. The mean sample age and VSMS score were respectively 61 months and 56.4. A normal score for children of that age is 56.5 . Thus, unlike in the PPVT test any deficiency in the social maturation of Head Start Children did not manifest itself in the scores. This may be due in part to the more subjective nature of the test.

As with the PPVT, the normal rate of change over a six month period varies with age--again with more rapid changes for younger children. At four years of age the normal pattern suggests a three point change each half year while at five years the change is 2½ points per half year. In contrast to the PPVT estimates, the pure age effects estimated in the VSMS regression--a 1.5 point gain for a six month period--are considerably less than this.[4] The estimate for weeks of exposure to Head Start indicates a 1.8 point change in a six month period. Thus the combined effects of age and weeks of exposure in the sample indicate a rate of development from 10-30% greater than the expected normal rate of development.

CONCLUSION

According to both the basic regression in Table III and the supplementary regression in Table V, the combined effect of age and exposure to Head Start produces an average increment of more than 5 points in the PPVT score over a 6-month period. This is around 150% of the mean 6-month increment in the general population of 5-year olds. At this rate around two years of Head Start training would be needed to overcome the average deficiency observed for children in the sample analyzed here. Of course there would be merit in a program that closed only half the gap; or in one which over-compensated and provided a real "head start."

[4] It must be remembered, however, that these estimates are based on cross-sectional differences and moreover, the lack of representativeness of the sample supports a suspicion that the coefficient for age has a downward bias.

This estimate of the Head Start's effect on PPVT scores is, however, very imprecise--a 95% confidence interval reaches very spectacular heights on the one hand, and includes negative values on the other. A balanced summary statement would perhaps indicate that a respectable gain was found by the analysis but that the evidence was too weak to disprove the contentions of doubters or enthusiasts.

For the social maturity variable, a more acceptable degree of precision was attained, but still the interval estimate was very broad. Here the combined effect of age and Head Start exposure was around 3½ points per six-month period, 35-40% greater than the change in the general population of 5-year olds. Interestingly, the sample analyzed here did not show any deficiency, on the average, in their rating on the VSMS relative to the general population. One would expect that two years of Head Start would provide this sample with several points of initial advantage over the average first grades.

The total cost per child of a two-year program, which would close the average gap in PPVT scores (8-10 points) and provide a 3-4 point advantage in terms of the VSMS, would be around $2250 taking an average of the '66 and '67 costs. The costs more directly related to the educational gains would account for nearly 70% of this total.

The next question is whether there is a less costly means of securing the same gains. Clearly, data about alternatives are needed for answering that. A closely related question is how much of the total cost should be allocated to the non-educational benefits of Head Start which have been ignored in this study. Some assumptions about prices to be attached to the medical, nutritional and other assorted benefits of Head Start are required in order to make a fair comparison with programs that provide more or fewer of such gains.

With the caution appropriate for generalizing from a relatively crude analysis of data which yield only weak evidence, this investigation has provided an initial basis for comparing Head Start with alternatives. There is a need for more and better empirical work of this kind, and it is hoped that the analytical methods used here can serve at least as a point of departure for others working on evaluative efforts.

Appendix

This Appendix contains four more supplementary regressions in addition to the eight sets of regressions mentioned in the text. Log linear forms were substituted for Weeks and Age in separate supplementary regressions to see if they significantly improved the fit. However, the R^2 in each case was not changed by more than 0.8%. One reason for using a log linear form for Weeks was to explore a different functional pattern in which Weeks and test scores are related. This would have some bearing on the optimum length for a Head Start program. If it were found that Head Start does not have much initial effect, but that there is some critical time period beyond which the rate of increase, or even the absolute increase, in the test scores improved, then programs might be extended somewhat beyond that critical time period. If, on the other hand, scores increased appreciably initially, but the rate of increase declined rapidly, it might only pay to have shorter programs, such as summer programs immediately preceding the academic year in which the child enters school. At this point we have made no interpretation of the results but have only presented them in the following tables. In each case the column heading gives the relevant dependent variable.

Appendix

Table A-1

REGRESSIONS USING LOGARITHMIC TRANSFORMATIONS FOR WEEKS AND AGE

	PPVT		VSMS		PPVT		VSMS	
Independent Variable	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
Constant term	2.16	6.08	36.337	3.707	-99.77	13.10	-10.783	7.968
1 Weeks					0.06	0.06	0.066	0.036
2 Hours	-0.06	0.33	0.226	0.203	0.01	0.33	0.253	0.202
3 Prev. PPVT	-1.05	1.47	-1.661	0.896	-1.04	1.46	-1.614	0.887
4 Prev. Train.	-0.64	0.79	0.537	0.479	-0.61	0.78	0.253	0.476
5 Age	0.57	0.05	0.255	0.031				
6a R_n	-5.75	0.89	0.386	0.540	-5.87	0.88	0.336	0.536
6b R_o	-2.08	3.37	-0.262	2.054	-1.89	3.36	-0.145	2.041
7 Sex	-1.42	0.74	0.961	0.451	-1.43	0.74	0.941	0.448
8 Father	1.76	1.15	1.768	0.703	1.69	1.15	1.710	0.698
9 Mother	-4.95	3.03	-4.280	1.845	-4.70	3.01	-4.143	1.833
10 Family Size	-0.05	0.18	0.004	0.112			-0.006	0.112
11 Siblings	1.09	1.74	0.929	1.060	0.92	1.73	0.872	1.052
12 $Ed._m$	1.41	0.34	0.144	0.209	1.40	0.34	0.144	0.208
13 Income	0.01	0.25	0.226	0.152	0.01	0.25	0.221	0.151
14 Welfare	-1.17	1.06	-0.716	0.646	-1.17	1.05	-0.696	0.641
15 Per Room	-0.97	0.43	-0.343	0.263	-0.96	0.43	-0.336	0.261
16 Durables	0.76	0.48	0.408	0.293	0.72	0.48	0.385	0.291
17 Newspaper	-1.38	0.98	0.526	0.597	-1.40	0.97	0.535	0.592
18 Water	-1.00	1.68	-3.662	1.025	-0.89	1.67	-3.605	1.018
19 Urban	-0.13	0.73	1.088	0.444	-0.10	0.72	-1.079	0.440
20a N.C.	0.60	1.25	1.523	0.761	0.60	1.24	1.547	0.757
20b S.	2.07	1.45	3.100	0.883	-1.47	1.44	1.411	0.874
20c W.	1.51	1.25	2.735	0.760	1.84	1.26	2.999	0.768
21 Lg. Weeks	0.75	1.00	0.780	0.612				
22 Lg. Age					33.47	2.98	15.508	1.812
R^2	0.354		0.242		0.360		0.253	

Table A-II

ADDITION OF VARIABLES BY GROUP

		BASIC REGRESSIONS				BASIC + GROUP I			
		PPVT		VSMS		PPVT		VSMS	
		Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
	Constant	44.12	1.61	57.97	0.91	16.16	3.69	39.40	2.22
1	Weeks	-0.05	0.06	-0.01	0.03	0.10	0.05	0.02	0.03
2	Hours	-0.61	0.34	-0.95	0.19	-0.29	0.30	-0.16	0.18
3	Prev. PPVT					-0.84	1.47	-1.44	0.88
4	Prev. Train.					-1.42	0.81	0.21	0.49
5	Age					0.46	0.05	0.27	0.03
6a	R_n					-5.15	0.87	0.31	0.53
6b	R_o					-0.50	3.47	0.78	2.09
7	Sex					-1.41	0.77	0.92	0.46
8	Father								
9	Mother								
10	Family Size								
11	Siblings								
12	$Ed._m$								
13	Income								
14	Welfare								
15	Per Room								
16	Durables								
17	Newspaper								
18	Water								
19	Urban								
20a	N.C.								
20b	S.								
20c	W.								

Table A-II (Cont.)

ADDITION OF VARIABLES BY GROUP

		BASIC + GROUP II				BASIC + GROUP III			
		PPVT		VSMS		PPVT		VSMS	
		Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
	Constant	44.89	2.99	59.76	1.68	43.71	2.98	57.86	1.69
1	Weeks	-0.06	0.06	-0.01	0.03	-0.08	0.06	-0.02	0.03
2	Hours	-0.64	0.34	-0.34	0.19	-0.61	0.34	-0.35	0.19
3	Prev. PPVT								
4	Prev. Train.								
5	Age								
6a	R_n								
6b	R_o								
7	Sex								
8	Father	-1.60	1.07	0.25	0.60				
9	Mother	-6.40	3.47	-5.26	1.96				
10	Family Size	-0.28	0.37	-0.16	0.21				
11	Siblings	0.09	0.42	0.11	0.23				
12	$Ed._m$	0.35	0.37	-0.22	0.21				
13	Income					-0.09	0.28	0.01	0.16
14	Welfare					-2.13	1.07	-0.99	0.60
15	Per Room					-0.76	0.44	-0.23	0.25
16	Durables					1.24	0.56	0.30	0.32
17	Newspaper					-0.96	1.13	0.99	0.64
18	Water					0.38	1.96	-1.89	1.11
19	Urban								
20a	N.C.								
20b	S.								
20c	W.								

Table A-II (Cont.)

ADDITION OF VARIABLES BY GROUP

		BASIC + GROUP IV				BASIC + GROUPS I AND II			
		PPVT		VSMS		PPVT		VSMS	
		Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
	Constant	36.86	3.77	53.94	2.06	4.00	4.41	37.31	2.73
1	Weeks	0.06	0.07	0.09	0.04	0.07	0.05	0.02	0.03
2	Hours	0.13	0.39	0.14	0.22	-0.26	0.29	-0.14	0.18
3	Prev. PPVT					-0.26	1.45	-1.17	0.90
4	Prev. Train.					-0.88	0.79	0.26	0.49
5	Age					0.52	0.05	0.28	0.03
6a	R_n					-6.20	0.87	0.03	0.54
6b	R_o					-2.37	3.38	0.30	2.09
7	Sex					-1.44	0.75	0.90	0.46
8	Father					0.86	0.92	0.85	0.57
9	Mother					-3.75	2.94	-3.79	1.82
10	Family Size					0.04	0.31	0.08	0.19
11	Siblings					-0.42	0.35	0.03	0.22
12	$Ed._m$					1.98	0.33	0.38	0.20
13	Income								
14	Welfare								
15	Per Room								
16	Durables								
17	Newspaper								
18	Water								
19	Urban	-0.44	0.86	-0.91	0.47				
20a	N.C.	4.14	1.44	2.45	0.79				
20b	S.	5.54	1.49	5.72	0.81				
20c	W.	5.36	1.45	3.91	0.79				

Table A-II (Cont.)

ADDITION OF VARIABLES BY GROUP

		BASIC + GROUPS I AND III				BASIC + GROUPS I AND IV			
		PPVT		VSMS		PPVT		VSMS	
		Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
	Constant	14.50	4.33	37.35	2.64	9.37	4.62	38.71	2.79
1	Weeks	0.06	0.05	0.14	0.33	0.10	0.06	0.08	0.04
2	Hours	-0.23	0.29	-0.11	0.18	-0.11	0.34	0.17	0.21
3	Prev. PPVT	-0.79	1.45	-1.80	0.89	-1.98	1.47	-1.69	0.89
4	Prev. Train.	-0.95	0.80	0.22	0.49	-1.27	0.80	0.24	0.48
5	Age	0.48	0.05	0.29	0.03	0.54	0.05	0.25	0.03
6a	R_n	-5.42	0.87	0.37	0.53	-4.68	0.88	0.63	0.53
6b	R_o	-1.20	3.41	0.66	2.08	0.17	3.43	0.72	2.07
7	Sex	-1.28	0.75	1.06	0.46	-1.51	0.76	0.85	0.46
8	Father								
9	Mother								
10	Family Size								
11	Siblings								
12	$Ed._m$								
13	Income	0.12	0.24	0.11	0.14				
14	Welfare	-0.16	0.92	0.00	0.56				
15	Per Room	-1.25	0.37	-0.37	0.23				
16	Durables	1.14	0.48	0.49	0.29				
17	Newspaper	-1.95	0.97	0.93	0.59				
18	Water	-2.22	1.67	-3.10	1.02				
19	Urban					0.19	0.74	-0.81	0.45
20a	N.C.					0.56	1.26	1.66	0.76
20b	S.					-3.47	1.42	2.39	0.86
20c	W.					2.11	1.27	3.15	0.77