A Follow-Up Study of Intelligence Changes in Children who Participated in Project Headstart*

by

David A. Waller and C. Keith Conners, Ph.D.

Johns Hopkins University School of Medicine

*This study was supported by funds from the Office of Economic Opportunity (Headstart Contract #510)

Abstract

Eighty-three children who had participated in a six-week Operation Headstart program and 83 matched controls who had not attended the program were retested nine months after the completion of the program. Earlier advantages of the Headstart children on the Peabody Picture Vocabulary Test were maintained by the Headstart children, their present scores still being significantly higher than those of their controls. Differences on the Draw-A-Person Test, though formerly significant, were now found to be non-significant. Since the intervening educational experience for Headstart and controls was identical, both having the same teachers and for the most part the same classes, it is concluded that Operation Headstart has provided the children with a relatively enduring gain in "verbal intelligence." It is noted, however, that a leveling effect appears to be taking place, with controls beginning to catch up.

In an evaluative study of the 1965 summer Headstart program in Baltimore (1), Eisenberg and Conners demonstrated significant gains in Peabody Picture Vocabulary Test (PPVT) and Draw-A-Person (DAP) raw scores in some 400 children who were completing Headstart training. In September, one month later, when the Headstart children enrolled in kindergarten, they were re-tested and found to have made further gains. At that time a control group was selected and matched with the Headstart group in age, sex, and race. They were children from the same neighborhoods as the Headstart children, but who had not been enrolled in the pre-school program. When the control group was tested in September, the scores were similar in mean and distribution to the Headstart group in June, but significantly lower than Headstart in September. Headstart children, therefore, entered elementary school significantly advanced with respect to those classmates who had not had the benefit of the program. The purpose of the present study was to determine whether in May, after nine months of school, any differences still existed.

Method

Sample. Considerations of time and staff made it necessary to select from the original sample a smaller group for follow-up testing, and the first concern was that this smaller group be representative. Eighty-three subjects were selected, together with their 83 previously matched controls, or about 22% of the initial sample of 382. The subjects came from 12 of the 16 summer Headstart centers operated by the public school system. Ten per cent of the original sample consisted of children enrolled in a church nursery Headstart program, but

because of the greater difficulties in following up these children, they were not included in the present sample. However, the previous study had revealed no differences between the gains made by children enrolled in the public and church operated programs. Ages in the previous (September) study ranged from 4 years, 1 month to 7 years, 5 months, the mean being 5 years, 2 months. In the present study the mean was 5 years, 11 months, with a range of 5 years, 4 months to 7 years, 4 months.

In other respects the samples closely corresponded. Both consisted of 54% boys, 46% girls. In the September study, 89% of the sample were Negro, the remainder, white; in the present study, 92% were Negro. That the present sample is in fact representative can best be demonstrated by comparing the previous test scores of the present sample with those of the larger group from which it was selected (Tables 1 and 2). Furthermore, when the original and present Headstart

---

Tables 1 and 2 about here

---

samples are divided into quartiles on the basis of their initial (June) scores, the changes from the beginning of Headstart to September are seen to correspond closely (Table 3).

Table 3 about here

A second major desideratum in selecting subjects was that the Headstart child and his control should have received the same, or nearly the same, classroom experience during the school year. To this end,

46 of the 83 subjects chosen were children whose controls had been in the same class. In order to increase the size of the sample, an additional 37 subjects were selected whose controls were in the same grade and had the same teacher, but at a different time during the day. Of these, 26 subjects were enrolled in morning kindergarten, their controls in afternoon sessions. The remaining 11 subjects were in afternoon kindergartens, the aim being to balance any effect of time of day on learning.

Materials and Procedure. The tests employed were Form B of the PPVT and the DAP. The Red Cross volunteers who administered the tests were chosen from the group that had participated in the previous study; all volunteers attended a brief re-orientation session to insure standardization of testing procedure. They were assigned at random to the children they were to test and were not informed whether the children were subjects or controls. No changes in the administration or scoring of the tests were introduced in the present study.

Results

PPVT. The results on the PPVT may be summarized as follows (Table 4):

Table 4 about here

---

1. Both Headstart subjects and their controls showed significant increases in raw scores since the September testing.
2. The average gain in the control group exceeded that in the Headstart group, the gains being 11 points and 8 points respectively.

3. Despite greater gains in the control group since September, the Headstart group remained significantly higher ($p < .05$) in PPVT raw scores, the means for the Headstart and control groups being 47.7 and 44.5, respectively.

When the two groups were divided into quartiles on the basis of their September scores, the greater increases were seen to have occurred in the lower quartiles in both groups (Table 5).

Table 5 about here

DAP (Table 6)

Table 6 about here

---

1. Both Headstart and control groups made significant gains.
2. Controls and Headstart gained about the same, the gains being 6 and 5 points, respectively; the difference was not significant.
3. There was no significant difference between the mean scores of the two groups, which were 14.9 and 14.6 for the Headstart and control groups, respectively.

Correlation. The correlation between PPVT and DAP raw scores in the present study was 0.36 for the Headstart group, 0.26 for the controls. The difference was not significant.

Discussion

Since the Headstart and control groups did not differ in their initial raw score distribution on the PPVT, and since they had been exposed to essentially the same educational environment since September, it would appear that the superiority demonstrated by the Headstart group in the present study on the PPVT could be attributed to the impact of the pre-school Headstart experience the previous summer. Factors which must be considered, however, as alternative explanations, include familiarity with test-taking, as well as familiarity with the test itself.

Although it might be postulated that general experience in test-taking accounted for a part of the superiority in IQ scores observed in the Headstart children, studies by Gray with Headstart children indicated no test-retest changes (2). Furthermore, the control group had been familiarized with taking the PPVT, and situations similar to, if not identical with, those involving the administration of the PPVT occur repeatedly during the school year. This experience would tend to minimize any advantage the Headstart group might enjoy as far as security in a test-taking situation is concerned. Of greater importance is the fact that whereas the control group was taking Form B of the PPVT for the first time, the Headstart group had been administered this form previously, at the conclusion of the Headstart program in August. Two factors are particularly relevant in this discussion:

1. Ten months elapsed between the two testings, including 9 months of school experience.
2. Taking the PPVT is not a learning experience, with respect to particular answers. The subject is never told whether

the picture he has selected does or does not correspond correctly to the word he has heard.

For these reasons it seems unlikely that the previous Headstart exposure to the PPVT influenced significantly the present results. Rather, the results suggest that the initial gain with which Headstart children began school permitted them to achieve levels of performance higher than their classmates during the school year.

Although Headstart children were significantly higher in performance on the PPVT in the present testing, the margin separating the two groups in September had narrowed; for while both groups showed significant increases during the school year, those of the control group were significantly greater (Figures 1 and 2). In hypothesizing

---

Figures 1 and 2 about here

---

a true leveling effect in two groups differing in mean IQ scores, consideration must of course be given to statistical regression. Yet it must be borne in mind that the scores of the two groups prior to any training closely corresponded in mean and distribution; and in the present study, sampling was from the entire distribution of each group. Hence, chance reversions to the original mean should not affect the relative position of the two means in the current testing, even though the difference between them was significant at the end of the Headstart program. A leveling effect should operate within the Headstart and control groups as well, and the variability in IQ should decrease (3). The fact that the standard deviations of both groups were lower in May than in September (Table 4) lends support to this interpretation.

Among the factors which could explain the larger gains made by children with lower initial scores are the following:

1. A ceiling effect, in which initially high scores reduce the possibility for gain.
2. A focus-of-instruction effect, instruction being aimed at the below-average student.
3. The test may have measured gains of more than one kind, suggesting the possibility that those with low initial scores made gains primarily in one area, more capable students making gains in another and presumably more difficult area.

It is difficult to assess the relative weights these considerations may have in the present study. The fact that the means for both groups in May were still well below the mean for middle class children tends to rule out a ceiling effect. Perhaps the most plausible explanation is that the instruction during the school year was non-specific in nature, with the result that only those children who were grossly deprived profited from it. The lower mean score of the control group in September could therefore account for its relatively more rapid gain during the year.

It has been reported that there was no significant difference between the performance of the two groups on the DAP, although both made significant gains over the September score. The advantage enjoyed by the Headstart group upon entering school in September, as measured by the DAP, was not as marked as that measured by the PPVT; and, perhaps not surprisingly, it was more easily washed out. Previous studies have demonstrated that specific training in art does not significantly

affect the DAP score (4). An obvious explanation for the fact that the results differed on the two tests, and that the correlation between the two tests was relatively low, concerns the essential difference in the qualities being measured. Whereas the PPVT is described as measuring primarily receptive language (5), the DAP is thought to mirror concept formation, or intellectual maturity, as well as perception (6). It appears, therefore, that the gains made by the Headstart group in the latter areas were less significant.

In a 1942 study of the relationship between language ability and intelligence (7), Dawe applied an educational program stressing factors related to superior language development to an experimental group of 11 pre-school and kindergarten children living in an orphanage. They were matched with a control group on the basis of school group, sex, chronological age, mental age, and IQ. The average amount of extra training received by each of the subjects was 50 hours. The experimental group increased an average of 14 IQ points on the Binet scale, while the controls showed no change. Moreover, the extent of IQ change was correlated with the per cent of time spent in individual training, which consisted, among other activities, of the presentation, definition, and discussion of new words. The present study differs, first in that the training was not so specifically related to the skills involved in taking a test; and secondly, in that gains were demonstrated to have survived at least 10 months after the training experience had concluded. Both factors constitute a convincing argument for the worth of a pre-school educational program, and against the notion of a fixed IQ.

While it is encouraging to demonstrate that objective benefits have been persistent in children who were enrolled in the summer 1965 Headstart program, despite nine months of an intervening educational situation lacking the advantages afforded by the summer project, optimism must be tempered by the realization that the relative superiority of the Headstart group as measured by the DAP has been erased, even while that reflected in the PPVT is disappearing. One may safely predict that without an attempt to maintain the benefits acquired from such a program, the opportunity for instituting significant and enduring changes will not be realized; and the promise of the headstart afforded these children will go unfulfilled.

References

1. Eisenberg, L. and Conners, C. K. The effect of Headstart on developmental processes. Presented at 1966 Joseph P. Kennedy Jr. Foundation Scientific Symposium on Mental Retardation, Boston, April 1966.

2. Gray, Susan. Personal communication.

3. Hunt, J. M. Intelligence and Experience. New York: The Ronald Press Company, 1961, p. 31.

4. Harris, D. B. Children's drawings as measures of intellectual maturity. New York: Harcourt, Brace & World, 1963, pp. 92-95.

5. Dunn, Lloyd M. Peabody picture vocabulary test; An expanded manual. Minneapolis, American Guidance Service, Inc., 1965.

6. Harris, D. B. Children's drawings as measures of intellectual maturity. New York: Harcourt, Brace & World, 1963, pp. 275.

7. Dawe, H. C. A study of the effect of an educational program upon language development and related mental functions in young children. J. exp. Education, 11: 200-209, 1942.

Acknowledgments

The authors wish to acknowledge with gratitude the assistance of Edith V. Walker, Catherine Brunner and Elaine Nolan of the Baltimore City Public Schools in arranging the study;

Red Cross volunteers who administered the tests;

Ann Goldman who scored the DAP;

Leon Eisenberg for his valuable suggestions for the manuscript and design of the study.

Table 1

Mean PPVT raw scores from previous testing of original and present samples

	Control	Headstart	
	C_1	H_1	H_3
Original Sample			
Mean	33.6	32.6	39.7
S.D.	11.7	12.3	11.3
(N)	(402)	(424)	(402)
Present Sample			
Mean	33.5	32.5	39.4
S.D.	11.0	13.1	12.3
(N)	(83)	(75)	(83)

C_1 = Controls in September, prior to any schooling

H_1 = Headstart in June, prior to program

H_3 = Headstart in September

Table 2

Mean DAP raw scores from previous testing of original and present samples

	Control	Headstart	
	C_1	H_1	H_3
Original Sample			
Mean	8.9	7.7	9.8
S.D.	5.0	4.8	4.4
(N)	(420)	(500)	(435)
Present Sample			
Mean	8.8	7.8	10.0
S.D.	5.0	4.5	4.5
(N)	(81)	(82)	(83)

C_1 = Controls in September, prior to any schooling

H_1 = Headstart in June, prior to program

H_3 = Headstart in September

Table 3

Mean change on PPVT from June to September for original and present Headstart samples arranged by quartile

	Quartile			
	1	2	3	4
Mean change for original Headstart sample ($H_3 - H_1$) (N = 382)	13.6	10.0	5.2	0.8
Mean change for present Headstart sample ($H_3 - H_1$) (N = 83)	11.8	9.4	4.7	1.4

Table 4

Mean PPVT raw scores from September and current (May) testing of present Headstart and control samples

	Control		Headstart	
	C_1	C_2	H_3	H_4
Mean	33.5	44.5	39.4	47.7
S.D.	11.0	8.9	12.3	10.8
(N)	(83)	(83)	(83)	(83)

C_1 = Controls in September

C_2 = Controls in May

H_3 = Headstart in September

H_4 = Headstart in May

Table 5

Mean change on PPVT from September to May for present Headstart and control samples arranged by quartile

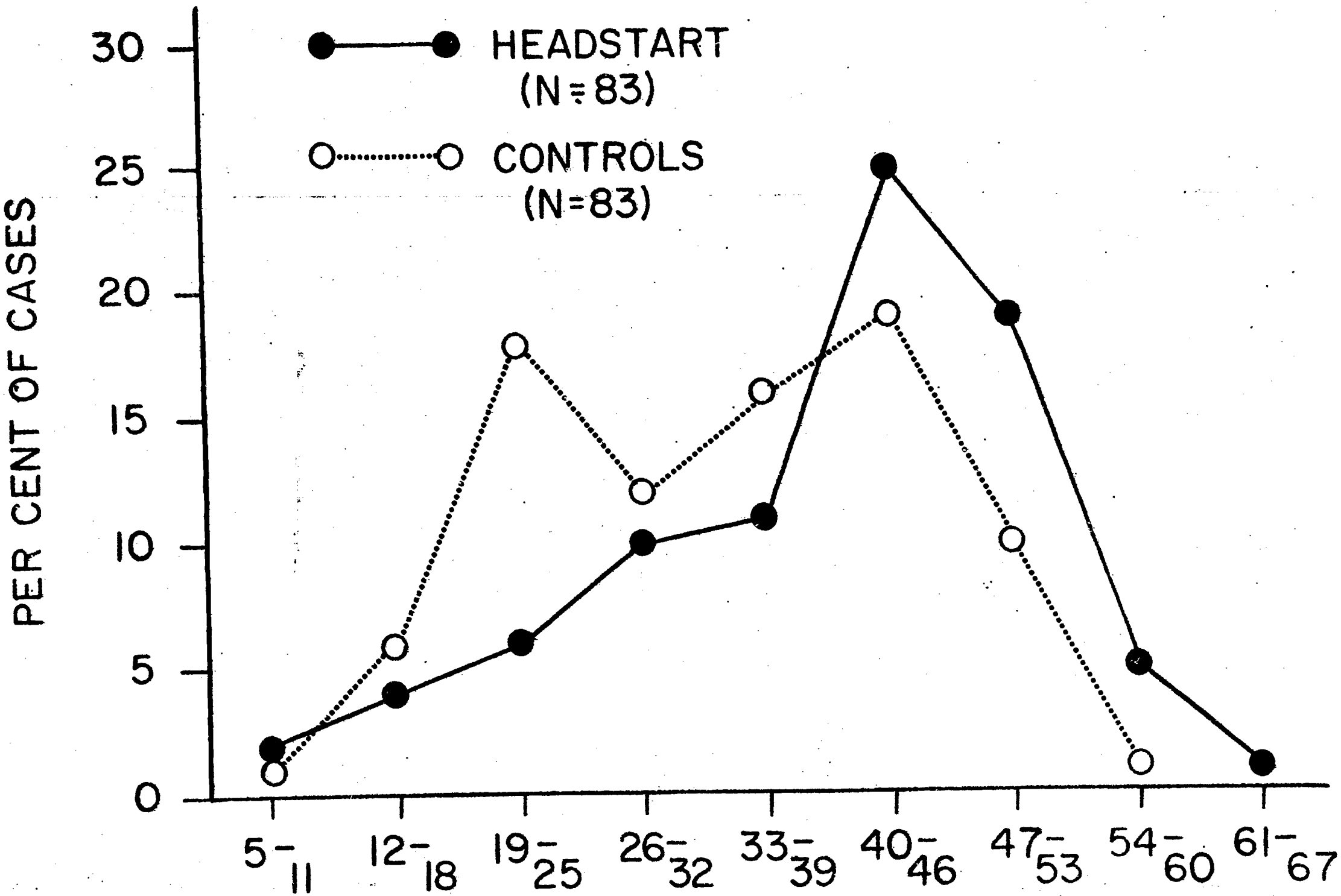
	Quartile			
	1	2	3	4
Mean change for present Headstart sample ($H_4 - H_3$) (N = 83)	16.0	6.9	5.0	4.4
Mean change for present control sample ($C_2 - C_1$) (N = 83)	17.1	15.6	9.6	2.8

Table 6

Mean DAP raw scores from September and current (May) testing of present Headstart and control samples

	Control		Headstart	
	C_1	C_2	H_3	H_4
Mean	8.8	14.6	10.0	14.9
S.D.	5.0	4.6	4.5	4.3
(N)	(81)	(83)	(83)	(83)

DISTRIBUTION OF PPVT RAW SCORES IN SEPTEMBER FOR PRESENT HEADSTART AND CONTROL SAMPLES



DISTRIBUTION OF PPVT RAW SCORES IN MAY FOR PRESENT HEADSTART AND CONTROL SAMPLES