DOCUMENT RESUME

ED 047 076 UD 011 284

AUTHOR Walberg, Herbert J.
TITLE An Evaluation of an Urban-Suburban School Bussing Program: Student Achievement and Perception of Class Learning Environments.
PUB DATE Feb 71
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, N.Y., February 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Academic Achievement, *Bus Transportation, *Integration Effects, *Integration Studies, Program Evaluation, *School Integration, Social Environment
IDENTIFIERS METCO, *Metropolitan Council on Educational Opportunity

ABSTRACT

This report is part of the evaluation of a program--Metropolitan Council on Educational Opportunity (METCO)--for the academic year 1968-69, which provides screening, placement, and busing services for Negro children from predominantly Negro schools in Boston to predominantly white schools in the surrounding suburbs. In this evaluation, METCO and non-random control children were tested close to the beginning and the end of the academic year, a unique feature of this evaluation being the use of siblings of the bused children as the control group; each control child selected was matched as closely as possible on age. With the exception that METCO children gained significantly less than the siblings on mathematics achievement at grade 5-6, there are no significant differences in performance between the two groups from grade 2-12. On a measure of the social environment of the classroom given at grades 3-4 and 5-6, the METCO children perceived their classes as more satisfying; METCO children in grades 5-6 also saw their classes as less difficult and competitive, and as having less friction. The evaluation concludes that school busing programs are a small step in the right direction, but may be doing too little too late. (Author/JW)

An Evaluation of an Urban-Suburban School Bussing Program: Student Achievement and Perception of Class Learning Environments*

Herbert J. Walberg
Harvard University
Now University of Illinois, Chicago

This report is part of the evaluation of the Metropolitan Council on Educational Opportunity (METCO) program for academic year 1968-69. The METCO program provides screening, placement, and bussing services for Negro children from predominantly Negro schools in Boston to predominantly white schools in the surrounding suburbs. Participation in the program is voluntary; and children with measured IQ's below 80 and those with serious emotional problems are excluded.

Matthai (1968) criticized a prior evaluation report on METCO (Archibald, 1967) and evaluations of school bussing programs in 10 other cities on two grounds:

1) Compared to non-participant groups, the possibly higher motivation of the students themselves may be the causes of success rather than the bussing program itself. Thus many findings must be held inconclusive in this regard.

2) The failure to include a control group of comparable non-participating children means that the analysis must be restricted to a) before-and-after comparisonson the only bussed children if both pretest and posttest scores are available and b) comparisons with "national norms" if only posttest scores are available. Comparisons of this kind are not very useful for evaluation since there is no evidence on how a comparable group of children who were not bussed perform.

The first problem can be solved 1) by taking a random sample of all children in the city and randomly assigning them to bussed and non-bussed groups or 2) by randomly assigning volunteer children to the two groups.

*Paper presented at AERA Meeting, New York, N.Y., February 1971

The first solution is nearly impossible because of political, administrative, and parental objections. The second solution is undesirable because children in the non-bussed group must be turned away even though they volunteered. Consequently, the second solution creates a negative image of the program in the minds of those who are turned down, and, worse yet, the bussed children may feel selected and elite. Hence, they may perform better during the year, not because of the program, but because they were selected.

Since these solutions appear unworkable, quasi-experimental comparisons of growth of non-randomly assigned groups have been proposed for field research in education (Campbell and Stanley, 1963). For example, in the present study, the METCO children and non-random control children were tested near the beginning and near the end of the academic year. This procedure allows comparisons of the relative progress during the year. However, since the sample is non-random, statistical inferences cannot be drawn beyond the sample.

A unique feature of this evaluation is the use of siblings of the bussed children as the control group. The control child selected for each METCO child was matched as closely as possible on age. This design feature by no means guarantees the equating of the groups since there may be bias in the family's choice of the child to be bussed; for example, the favorite child or the child wanting to go to school with white children may be more likely to be sent. On the other hand, siblings are likely to be exposed to similar family and neighborhood environments.

Before turning to the procedures and results of the evaluation, it may be worthwhile noting a few difficulties characteristic of research in this area that may be relevant to the present evaluation. Research in education has revealed little evidence that differing school characteristics, curricula, teacher characteristics, and instructional methods and media make for significant differences in the rate of learning. A current review of reviews of educational research (Stephens, 1967) showed that the factors that are often said to give suburban schools advantages over urban schools actually do not make for increased rates of learning. These factors include the administrative organization of the school, school size and lavishness of facilities, the presence of specialized teachers and guidance counselors, teacher education, knowledge, and experience, and class size and ability grouping. There may be several reasons for the apparent stability of learning rates despite what appear to be promising educational interventions. First, even the most ardent environmentalists acknowledge the relative importance of genetic factors in the determination of intelligence. Roughly 70 percent (the mean of estimates reviewed by Bloom, 1964) of the variance in intelligence may be attributed to heredity. Second, Bloom's review of longitudinal studies also showed the importance of the child's early environment, particularly during the first five years, in predicting later achievements. Late environmental interventions have less impact unless they are extreme. Third, interventions of less than a year's duration that only change some unspecified aspects of some parts of the school child's environment for some

fraction of his daily life are unlikely to produce dramatic changes in rates of general learning. When these considerations are weighed against the possible benefits of interventions, the overall evidence often supports the hypothesis of no significant differences between experimental and control groups. Moreover, intervention may also have unanticipated consequences (such as waiting for busses on cold corners during the winter and long rides through heavy traffic as in the METCO program) that may vitiate its possible benefits. Thus the null hypothesis of no differences between the academic performance in bussed and non-bussed groups seems warranted for the present research.

Method

Sample

The parents of all the METCO children were requested to bring in their children for testing in October of 1968 and again in May of 1969. As mentioned earlier, a control child from each METCO family was selected closest in age to the METCO child. The letter brought out the need for cooperation in testing each METCO child and the sibling for an adequate evaluation of the program. It was promised that a summary of the report would be made available to the parents for their information and that the scores of their children would be reported to them individually and confidentially. These promises were kept.

Instruments

Since they are well-regarded nationally-standardized instruments

measuring general achievement in various school subjects, the Metropolitan Tests (Durost, 1964) were selected.

Form A was administered in the fall, and Form B was administered in the spring. The sub-tests given and their median split-half corrected reliabilities (reported in the manuals) are as follows:

Grades 3-4	
Reading	.90
Arithmetic Problem Solving and Concepts	.92
Grades 5-6	
Reading	.90
Arithmetic Problem Solving and Concepts	.92
Grades 7-9	
Reading	.90
Arithmetic Problem Solving and Concepts	.91
Grades 10-12	
Reading	.82
Arithmetic Problem Solving and Concepts	.85

Reading and math sub-tests were chosen because they are of interest to parents and teachers because other school subjects often presuppose achievement in these basic areas. Raw scores were used in all analyses.

The My Class questionnaire was adapted for elementary school students from the Learning Environment Inventory (Walberg and Anderson, 1968; Walberg, 1969). It measures the students perception of the classroom group by requesting his agreement or disagreement (in 4 degrees) with 45 items describing his class. Nine items are averaged to score each

of the five scales. The scales, sample items, and scale reliabilities (Spearman-Brown boosted split-half on the present sample) are listed below:

Satisfaction	The pupils enjoy their schoolwork in my class.	.71
Friction	Children are always fighting with each other.	.73
Competitiveness	Some pupils always try to do their work better than the others.	.56
Difficulty	In our class the work is hard to do.	.61
Intimacy	All the pupils in my class like one another.	.57

Early research (cited above) with the high school form of the instrument showed that the scales can be predicted from the size and composition of the class, and that the scales are valid in predicting cognitive and affective learning with other relevant variables held constant statistically. For the present evaluation, the children in grades 3-4 and 5-6 were given the questionnaire during the spring testing day. They were asked to describe the class in which they had spent the most time during the school year. The intent was to compare the perceptions of class environments of bussed and non-bussed children.

Procedure

Negro school teachers and other qualified Negroes from both Boston and suburban public schools were recruited to administer the tests and questionnaires in October and again in May. At the training session a week before the first testing, the teachers became acquainted with the purpose of the research and learned how to administer the instruments according to the standardized instructions. They were also impressed with the need for both objectivity and rapport in dealing with the children. The administrators were asked to involve the

children in the research by asking them to be honest since the findings might suggest ways of helping other children in their schoolwork. During the examinations, psychologists and other qualified personnel circulated from room to room to check the testing conditions and to answer questions raised by the proctors. During the first testing, it was discovered that one test administrator had mixed up the instructions and mistimed the reading tests. All scores for children in this room were excluded from the analysis. Aside from this error, the test administration appeared to meet conventional standards. The elementary children were tested in a Boston elementary school, an adequate, clean building. However, the high school students were tested in a Boston technical school, an old, run-down, ill-cared-for building. The conditions of this building may have affected the motivation of both the METCO children and their siblings in taking the high school tests on both occasions.

Analysis

For the first main analysis, univariate statistics were computed for METCO children and the siblings for each test level. Inspection of these figures revealed no apparent departures from the normal distribution with respect to skewness and kurtosis. However, the number of cases on the sets of two groups differed widely and the standard deviations differed moderately. Hence, Welch T-tests, which make no assumptions of equal numbers or equal variances, were employed to test the significance of differences in the pretests, posttests and gains for the two-group comparisons. Also, as a preliminary test of interaction, these analyses were performed separately for boys and girls at each test level.

Inspection of the T-values calculated separately for boys and girls revealed no tendency toward interaction of sex and group except possibly for grades 3-4 and 5-6. Hence, on these groups, a more powerful and sensitive analysis --ordered, stepwise multiple regression with product terms -- was carried out. Some workers (Campbell and Stanley, 1963) believe that regression-adjusted gain scores are more accurate than raw differences between pretests and posttests. For this reason, each posttest, (reading and mathematics), was predicted by its corresponding pretest in the regression models. After this, in first regression model, the group term (a binary variable -- 1 or 0 bussed or non-bussed) was added to test the significance of difference between the two groups. Lastly, in the first regression model, the product of the pretest and group was added to test the interaction. The test of this term is formally equivalent to the linear interaction in analysis of variance and to a heterogeneity of linear regression in covariance (Ahlgren and Walberg, 1969).

The second regression model also employed the corresponding pretest as the first term, but added in succession the binary variables -- sex, group, and the product of sex and group. Thus the two models provide sensitive tests of the group-pretest and the group-sex interactions.

Results and Discussion

Table 1 shows the turnout rates for METCO and sibling groups at four test levels for the pretest and posttest administration. A few METCO children in grade 2 took posttests, but no siblings did. Therefore, this group was excluded from the analysis. The numbers of children taking tests at each level are probably large enough to detect any true differences statistically; however, the differences in turnout

rates, scores, test levels and groups cast doubt on the comparisons reported below. It can be noted that greater percentages of eligible elementary school children took the tests than did those in junior and senior high school. Even more serious is the fact that higher percentages of the METCO children than siblings took the tests. It would be comforting to assume that the more highly motivated siblings turned out, thereby biasing the results against METCO and providing a more rigorous test of its effectiveness, but this assumption is probably unwarranted, and it is difficult to say in which direction the results are biased.

Inspection of Table 2 reveals that there are no significant ($p < .05$) differences in reading scores between the METCO children and the siblings on the pretests, posttests, and gains for all test levels. Table 3 contains two significant differences between the groups on the mathematics tests. At grade level 7-9 the METCO children scored significantly higher on the posttests than the siblings; however, the gain scores are not significantly higher for the METCO group, perhaps because the METCO children were slightly, but not significantly higher on the pretest. The other significant difference on the mathematics test shows that at grade level 5-6 the METCO children gained significantly less than did the siblings.

Table 4 shows the comparisons for grades 3-4 and 5-6 on the My Class Scales. METCO children at both grade levels perceived their classes as significantly more satisfying than did the siblings. Also,

METCO children in grades 5-6 saw their classes as less difficult and competitive and as having less friction.

(Although the original intent of the evaluation was to simply compare the achievement gains and class perceptions of METCO and sibling groups, it had been suggested that children participating in the program for the first, second, or third year might gain at different rates as compared with the siblings. However, F-tests across the four groups and T-test comparisons of each of the three year groups with the siblings showed no statistically significant differences with respect to year in program.)

It was mentioned earlier that T-tests on the responses of boys and girls separately revealed a slight possibility of a group-sex interaction only for grade levels 3-4 and 5-6 and that stepwise regression tests were computed to test this possibility and the futher possibility of group-pretest interaction. However, Table 5 reveals that this more powerful analysis merely confirms the main-effect, T-test comparisons and reveals no significant interactions: that is, with the corresponding pretests held constant, there is no significant tendency for boys to differ from girls in posttest reading or mathematics achievement as a result of the METCO program; nor is there any tendency for initially high achieving children to differ from the others on the posttests as a result of the program. The T-test differences in classroom perception between the groups are not strongly confirmed in a multivariate sense (Bock and Haggard, 1968): the groups do not differ on the five

scales collectively at the .05 level. However, the multivariate tests were significant at the .10 level; and the reader may or may not accept this error rate. The univariate regression tests on each scale separately confirmed the T-test differences described earlier.

Summary and Conclusion

With the exception that METCO children gained significantly less than the siblings on mathematics achievement at grade level 5-6, there are no significant differences in academic performance between the two groups from second through twelfth grade levels. Nor did sex or initial achievement interact with group. On a measure of the social environment of the classroom given at grades 3-4 and 5-6, the METCO children perceived their classes as more satisfying. METCO children in grades 5-6 also saw their classes as less difficult and competitive and as having less friction.

Methodological difficulties of field research in education obviously bear upon the present evaluation. The sample was not randomly drawn, nor were the children randomly assigned to groups. However, data collected over time on the bussed children and a group of siblings afforded about as much statistical control and comparability as feasible for this kind of research. While the sample sizes are probably high enough to detect true differences, the sampling of both groups is biased in unknown ways; for example, the more conscientious parents may have insured that their children attended both testing sessions. It may be too much to hope that such biases equally affected the turnout of the METCO children and siblings. Short of testing in the home or paying children to take tests (which also introduce methodological prob-

lems), the turnout may be as good as one can expect considering the effort that went into obtaining adequate sampling.

The results are disappointing, but perhaps predictable on the basis of the reviews of educational research referred to earlier. Nor can too much comfort be taken in finding that the bussed children find suburban classes more satisfying, for it is performance on achievement tests that makes for academic success, that in turn often opens the way to many careers and to higher socio-economic status in this increasingly meritocratic society. At the same time, it must be recognized that there are many factors such as aspiration, social awareness and integration, and creativity that may have been affected by the program. These factors are difficult to assess with psychometric tests and scales, and measurement of these factors has not been attempted in this evaluation. Hopefully, parallel evaluations being carried out on the METCO program by clinical psychologists, sociologists, and political scientists may reveal changes in these and other factors as a result of the program.

In conclusion, it is this writer's view that school bussing programs such as METCO and pre-school programs such as Head Start are a small step in the right direction, but they may be doing too little, too late. For only part of his day and part of his life, they bring the child into what may be a more stimulating academic and social environment of the suburban schools or urban enrichment centers. Moreover, they are addressed to a time in the child's life when physical and psychological growth rates are relatively stabilized. Bloom's (1964) review of longitudinal research strongly points to the child's

environment, particularly the home, during the first four years of life as most crucial for growth and later achievement. Thus, though perhaps unacceptable in present family, social, and political spheres, the continuous, intensive enrichment of infant and early childhood environments may be the most potent means of giving urban children from poor families an equal opportunity in school and in life. Short of environmental interventions at these ages, bussing and enrichment programs for pre-school and school age children should be continued and expanded vastly. Programs such as these may be less powerful than earlier interventions, but they seem to be the only hope for equalizing opportunities in the near future. Hopefully, continuing assessment of these programs may identify the factors that make for increased effectiveness.

References

Ahlgren, Andrew and Walberg, Herbert J. "Homogeneity of Regression Tests: Assumptions, Limitations and Alternatives," American Educational Research Journal, 1969, 6, 696-700.

Archibald, David K. Report on Change in Academic Achievement for a Sample of Elementary School Children: Progress Report on METCO (Public Law 89-10, Title III, Project #661405, U.S. Office of Education) Brookline, Mass., 1967, mimeo.

Bloom, Benjamin S. Stability and Change in Human Characteristics. New York: John Wiley, 1964.

Bock, R. Darrell and Haggard, Ernest R. "The Use of Multivariate Analysis of Variance in Behavioral Research," in Dean K. Whitla (Ed.) Handbook of Measurement and Assessment in the Behavioral Sciences. Reading, Mass.: Addison Wesley, 1968. Pp. 100-143.

Campbell, Donald T. and Stanley, Julian C. "Experimental and Quasi-Experimental Designs for Research on Teaching," in N.L. Gage (Ed.), Handbook for Research in Teaching. Chicago: Rand-McNally & Co., 1963. Pp. 171-246.

Durost, Walter N. Metropolitan Achievement Tests Manual, New York: Harcourt, Brace & World, Inc., 1964.

Matthai, Robert A. The Academic Performance of Negro Students: An Analysis of the Research Findings from Several Bussing Programs. Cambridge, Mass.: Harvard Graduate School of Education, unpublished doctoral qualifying paper, 1968, mimeo.

Stephens, J.M. The Process of Schooling: A Psychological Examination, New York: Holt, Rinehart, and Winston, Inc., 1967.

Walberg, Herbert J. "Predicting Class Learning: A Generalized Regression Approach to the Class as a Social System." American Educational Research Journal, 1969, 6, 529-542.

Walberg, Herbert J. and Anderson, Gary J. "Classroom Climate and Individual Learning," Journal of Educational Psychology, 59, pp. 414-419, 1968.

Table 1

Turnout Rates for Two Groups on
Two Occasions at Four Grade Levels

Grade	Total Eligible		% Tested in October		% Tested in May		% Tested on Both Occasions	
	METCO	Control	METCO	Control	METCO	Control	METCO	Control
-4	175	46	73.1	71.7	60.6	58.7	51.4	37.0
-6	116	98	78.4	51.0	62.9	54.1	52.6	29.6
-9	288	112	43.1	25.0	50.0	42.0	43.1	25.0
-12	158	96	45.6	14.6	54.4	20.8	45.6	14.6

Note: Actual numbers for each comparison are shown in Tables 2, 3, and 4.

Table 2

Univariate Statistics and Welch T-tests on Reading Scores

ade	Statistic	Pretest METCO	Pretest Control	T	Posttest METCO	Posttest Control	T	Gain Score METCO	Gain Score Control	T
-4	M	19.63	22.03	-1.43	21.24	22.52	-.65	1.80	2.00	-.15
	S	8.22	8.65		7.12	9.60		5.77	4.65	
	N	128	33		106	27		90	17	
-6	M	20.69	20.28	.36	24.53	23.70	.63	3.62	5.03	-1.40
	S	6.52	6.44		6.94	7.57		5.25	4.04	
	N	91	50		73	53		61	29	
-9	M	25.83	25.96	-.07	27.78	26.23	.99	2.13	1.46	.72
	S	7.14	10.15		7.02	9.97		5.13	4.25	
	N	124	28		144	47		124	28	
-12	M	19.26	15.93	1.54	20.76	18.85	1.18	1.22	3.29	-.83
	S	5.15	7.61		6.01	6.63		5.20	9.06	
	N	72	14		86	20		72	14	

te -- Minus signs are given on the T values to show the direction of the differences and to enable scanning of the table.

No comparison is significant (probability less than .05).

Table 3

Univariate Statistics and Welch T-tests on Mathematics Scores

Grade	Statistic	Pretest			Posttest			Gain Score		
		METCO	Control	T	METCO	Control	T	METCO	Control	T
3-4	M	8.59	11.00	-1.91	11.82	14.67	-1.43	3.54	4.29	-0.40
	S	3.35	6.72		6.81	9.74		4.79	7.46	
	N	128	33		106	27		90	17	
5-6	M	7.46	6.20	1.34	18.62	21.23	-1.86	10.47	14.79	-3.20**
	S	5.94	5.01		6.88	9.37		5.35	6.25	
	N	91	50		73	53		61	29	
7-9	M	16.95	15.79	.60	20.24	17.21	2.15*	3.31	2.04	1.48
	S	7.09	9.67		7.88	8.50		5.66	3.71	
	N	124	28		144	47		124	28	
10-12	M	9.96	10.36	-.68	11.93	12.25	-.24	1.99	3.00	-.83
	S	4.67	4.54		4.32	5.47		3.66	4.26	
	N	72	14		86	20		72	14	

Note -- T-tests significant at the .05 and .01 levels respectively are indicated with one and two asterisks.

Table 4

Univariate Statistics and Welch T-tests on Classroom Environment Scales

Environment Scale	Statistic	Grades 3-4 METCO	Grades 3-4 Control	Grades 3-4 T	Grades 5-6 METCO	Grades 5-6 Control	Grades 5-6 T
Satisfaction	M	1.35	1.50	-2.39*	1.32	1.45	-2.48*
	S	.24	.30		.27	.31	
	N	106	27		73	53	
Friction	M	1.50	1.39	1.77	1.52	1.39	2.46*
	S	.26	.30		.28	.28	
	N	106	27		73	53	
Competitiveness	M	1.35	1.34	.28	1.42	1.32	2.54*
	S	.22	.21		.23	.21	
	N	106	27		73	53	
Difficulty	M	1.59	1.57	.56	1.74	1.63	2.88**
	S	.20	.20		.17	.23	
	N	106	27	73.	73.	53	
Intimacy	M	1.28	1.37	-1.92	1.28	1.31	-.74
	S	.21	.21		.21	.19	
	N	106	27		73	53	

Note -- T-tests significant at the .05 and .01 levels respectively are indicated with one and two asterisks. Lower means denote higher scores on the scales.

Table 5

Chi-Square Approximation to Wilk's Lambda for Adding Successive Terms to Two Regression Models

Grade Level / Dependent Variables	Model 1			Model 2			
	Pretest	Group	Pretest-Group Interaction	Pretest	Sex	Group	Sex-Group Interaction
Grades 3-4							
Reading	74.67**	.28	2.42	74.67**	.01	.28	.14
Mathematics	80.23**	.03	.31	80.23**	.00	.03	1.39
My Class	.13	.80	3.67	--	2.29	9.27	2.82
Grades 5-6							
Reading	70.41**	1.09	.62	70.41**	4.83	1.09	.95
Mathematics	42.11**	10.80**	.20	42.11**	.01	10.80**	.24
My Class	5.53	10.10	2.54	--	7.56	10.10	5.04

Note -- The Chi-square values are univariate tests for Reading and Mathematics scores and multivariate tests on the five My Class scales; in Model 1 reading scores were used as the pretest in predicting the My Class scales. No pretest was employed to predict the My Class scales in Model 2. Terms contributing to the variance significantly at the .05 and .01 levels are indicated with one and two asterisks, respectively. The numbers of cases for grade levels 3-4 and 5-6 are 105 and 100, respectively. The multivariate tests for the group effect on the 5 My Class scales simultaneously were significant beyond the .10 level (see text).

Abstract

Despite public interest and controversy regarding bussing programs for disadvantaged urban children, political and logistical problems have limited past objective evaluations in large cities to pre- post-test comparisons on achievement without comparable control groups. Accordingly, a quasi-experiment was conducted in Boston on a population of 737 bussed children and 352 siblings matched on age. Although poor turnout rates introduced rival hypotheses, the bussed children gained about as much on reading and mathematics as their siblings although elementary school children rated the suburban classroom environments more satisfying. Some difficulties in evaluation of social intervention programs and the determination of social policy in education are discussed.