DOCUMENT RESUME

ED 239 000 UD 023 303

AUTHOR Armor, David J.
TITLE The Evidence on Desegregation and Black Achievement.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Mar 83
NOTE 54p.; Paper prepared for the National Institute of Education Panel on the Effects of School Desegregation. For related documents, see UD 023 302-308.
PUB TYPE Information Analyses (070) -- Viewpoints (120)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Gains; *Black Students; *Desegregation Effects; Elementary Secondary Education; *Mathematics Achievement; Meta Analysis; Outcomes of Education; Performance Factors; Program Effectiveness; *Reading Achievement; Research Methodology; Research Reports; School Desegregation

ABSTRACT

A brief survey of several comprehensive reviews of studies of the effects of desegregation on black achievement makes it clear that the experts are still in disagreement, and it is in the hope of finally bringing about a consensus on the subject that the National Institute of Education (NIE) has established a representative panel of experts to review the evidence and pass judgment on this issue. The 19 studies selected for review and analysis by the NIE panel did not include cross-sectional studies or longitudinal studies without a control group; all the selected studies fulfilled a quasi-experimental design, with pre- and post-tests as well as a segregated control group. Analysis of the studies in terms of each reseacher's own significant tests and the meta-analysis technique employed by the NIE panel shows no significant and consistent effects of desegregation on black achievement. There is virtually no effect for math achievement. For reading, the best that can be said is that only a handful of grade levels show substantial positive effects, while the large majority of grade levels show small and inconsistent effects that average out to almost zero. The fact that only a small fraction of these studies show substantial effects strongly suggest that factors other than desegregation are the real causes of the large achievement gains that are documented. (CMG)

THE EVIDENCE ON DESEGREGATION AND BLACK ACHIEVEMENT

David J. Armor, Ph.D.

Prepared for the National Institute of Education
Panel on the Effects of School desegregation

March 1983

PAGES 1 THROUGH 8 MISSING

2/9

The debate over the costs and benefits of school desegregation, particularly in its mandatory forms, continues unabated today, nearly 30 years after the fateful Brown decision by the U.S. Supreme Court. No issue has been more central to this debate than the question we address here: the impact of desegregation on black student achievement.

Indeed, it is remarkable that this question remains in controversy today, considering the extent of school desegregation over the past twenty years and especially given the mandatory methods imposed by the courts over the past fifteen years. One wonders how many courts have ordered busing, how many agencies have allocated time and money, and how many black parents have willingly sent their children to distant schools out of their neighborhoods, on the assumption that desegregation would yield academic benefits for black children.

Obviously, more is at stake in desegregation policy than the academic progress of students. Desegregation is a highly desireable social policy regardless of its educational benefits, and many educators and parents will and should seek it despite research findings. On the other hand, it is one matter to agree that school desegregation is a desireable policy and quite another to make it compulsory regardless of other considerations. The moral imperatives permitting coercion in social policy make it unlikely, in my opinion, that our courts would have abandoned the traditional neighborhood school policy in favor of mandatory busing without the belief that they were actually benefiting the education of black students. Why else would so many courts hear evidence, and so many legal journals publish treatises on this issue?

Aside from the legal importance of the achievement question, it does have immediate relevance to educational policy-makers, especially in this day of tight budgets. It is beyond dispute that we need programs to enhance minority achievement. The key question is, what kinds of programs? In recent years significant amounts of time and money have been devoted to improving racial balance in schools, justified in part by its supposed educational payoffs. Is this resource investment in fact yielding a fair return, in terms of improving minority achievement, or would other programs have greater impact? In other words, are racial balance activities cost-effective when compared to other available alternatives? If not, we should re-order our priorities and invest in programs that promise to work.

Finally, the issue of desegregation and black achievement should have more than a passing interest to parents of black children, who for years have borne the heaviest personal cost of desegregation by enduring long bus rides, seperation from familiar surroundings, and curtailment of extracurricular activities. It is quite likely that, over the long run, black parents support of busing for the purpose of desegregation would lessen if desegregation was found to have minimal impact on their children's rate of learning.

For all these reasons, the National Institute of Education must be commended for bringing together, for the first time, a representative panel of experts to review the evidence and pass judgement on this difficult but vital issue. At the same time, more than one observer will be surprised at the small number of studies (19 in all) meeting the minimal scientific standards established by the panel, and perhaps shocked that only <u>three</u> of these studies have been conducted within the past ten years, when school desegregation has been at its peak.* It is almost as though educational researchers and their funding agencies -- including NIE -- believe that the issue is settled, or no longer important. It is clearly an important question, and even a cursory review of the available literature shows that it is clearly unsettled. Hopefully, this panel will offer a consensus judgement that will finally settle the controversy.

Before turning to the studies selected for review by the NIE panel, I will comment briefly on several other comprehensive review efforts. To a large extent the approach taken by the panel culminates an evolutionary sequence that can be observed in the previous attempts to grasp the essential truths in this varied and complex literature.

* Different panelists, including myself, will take methodological exception to some of these studies.

PREVIOUS REVIEWS

Much of the early disagreement over the desegregation and achievement issue stemmed from reliance on a single study, or on a small number of studies where variation in results and conclusions might be expected (e.g., Armor, 1972 and 1973; Pettigrew, 1973). Yet disagreement persists even among the comprehensive reviews, all of which investigate many of the same studies.

The first review to encompass a large number of studies was carried out by Weinberg (1970). Like his most recent review, Weinberg covers a lot of studies but makes little or no attempt to select studies according to their methodological adequacy for causal inference (Weinberg, 1977). As we shall see, his conclusion that desegregation significantly benefits minority achievement was undoubtably affected by his failure to consider a study's scientific rigor.

The second comprehensive review by St. John (1975) made considerable progress over Weinberg. Not only was her study coverage broad, but she additionally classified studies according to the research design employed, allowing her to observe the relationship between methodology and the impact of desegregation. When St. John took design rigor into account, she reported that the evidence was mixed, preventing a firm conclusion about the benefit of desegregation for black achievement. A later review by Bradley and Bradley (1978) did not expand on the state of the art over St. John. They did conclude that methodological flaws impaired the entire group of studies, and that nothing could be decided.

A distinct advance was made in Krol's (1978) review, where he applied formal "meta analysis" to 55 studies, as that phrase has been used by Glass (1978) and others. The technique Krol used involved two critical steps that are lacking in previous reviews. First, studies were screened for minimal methodological adequacy (e.g., appropriate treatment condition and quantitative results) and coded as to a variety of conditions related to the type of research design and other study attributes. Second, achievement test results were converted to quantified standardized estimates by taking the ratio of test score means to their standard deviations. This allows estimates of the magnitude of segregation effects, as well as the impact of specific study characteristics on those effects.

Using this approach Krol concluded that the average effect of desegregation on black achievement is .16 standard deviations, which (depending on the type of achievement test) amounts to anywhere between 1½ to 3 months of progress during an academic year. However, this effect was not statistically significant, and the effect for that subset of studies with a valid control group was only .10, which again was not significant. The major limitation for the Krol study is that the number of studies was small, and no adjustment was made for control group selection bias; that is, for treatment-control differences prior to treatment. Moreover, the way he estimated effects for studies without control groups assumed that a control group would experience no gain. This is not a tenable assumption for achievement test data, where some academic growth is the norm for most students at least through the 10th grade.

The most recent large-scale review was carried out by Crain and Mahard in several stages (1982). The latest version of this review also uses the meta-analysis approach, with quantified effect estimates and study characteristics coded for some 93 studies. Although the number of studies is larger than in Krol's review, Crain and Mahard intentionally included studies with weaker design characteristics in order to test the impact of design flaws on desegregation effects. Their overall effect size mean is .065 standard deviations, which is both negligible and non-significant.

Crain and Mahard do find differential major effects for grade level, with an average effect size nearing .3 for students desegregated at the kindergarten or 1st grade level, but dropping off markedly to near 0 in the 2nd and higher grades. On the basis of this finding, they argue that desegregation can have a significant effect on black achievement, providing it starts in or before the 1st grade; it will have little or no effect on students starting desegregation in later grades. It is not clear from the study whether this effect occurs only at these early grade levels, or whether it is cumulative. In any event, there are some further methodological problems with this conclusion. It appears, for example, that none of the studies which have tested Kindergarten and 1st graders have been adjusted for possible selection bias, which continues to be a major problem in this field. We will take this issue up once again in our concluding section, after reviewing the NIE studies.

NIE STUDY PROCEDURES

It is clear from the foregoing review that there is still disagreement among the experts about the effect of desegregation on black achievement. The purpose of the NIE panel is to establish methodological guidelines for selection of studies, to review the studies so selected, and to decide what these studies say about the effect of desegregation on black achievement. I will comment briefly on these guidelines, leaving their major exposition in the capable hands of Dr. Wortman.

Study Selection Guidelines

The major reason for variations in conclusion of major reviewers is that they are looking at different sets of studies, which vary greatly as to their adequacy for making a causal inference. By establishing "minimum" standards for selecting studies, the NIE panel does not mean that the resulting set is "pure". Indeed, there may be no such studies in existence. The very nature of the process being studied prevents the ideal experiment, where one can eliminate all confounding factors but the factor being tested. It is believed, however, that studies selected according to these guidelines have the best chance for arriving at a decision about whether desegregation itself -- and not other factors -- was responsible for changes, if any, in black achievement.

For example, the guidelines exclude cross-sectional studies, because they do not allow determination of whether desegregated students have actually gained on the achievement test in question compared to segregated students, or whether differences simply reflect prior differences between segregated and desegregated students that persist over time.

Likewise, longitudinal (over-time) studies without a control group of some kind are also excluded since some academic growth can be expected of nearly all students during their school career, regardless of desegregation experiences. A segregated control group is necessary if one wishes to conclude that desegregated black students have gained or lost in comparison to black students who remained in segregated schools.

Thus, in addition to the usual requirements of quantifiability, relevance, and so forth, all selected studies fulfill a basic quasi-experimental design, with pre- and post-tests as well as a segregated control group (where segregation is defined as 50 percent or more black). We do not imply, however, that there are no further methodological problems. Only one of the studies selected is a randomized experiment and therefore the control group is not generally equivalent to the treatment group prior to the start of desegregation. Wortman's preliminary analysis shows that the correlation of pre-test and post-test effect sizes is .74. This condition raises a serious threat to causal inference, because -- just as in a cross-sectional study -- any observed differences between desegregated and segregated students after desegregation could simply reflect pre-existing differences between the treatment and control groups.

Fortunately, the selection criteria also require pre-test means to ensure that adjustments can be made to remove the pre-treatment effects. As we shall see, adjusting the control groups for initial differences has a significant impact on one's conclusions from these 19 studies.

I disagree somewhat with two of the guideline provisions. First, the adjustment method to be described in the next section is not infallible and is itself based on a number of assumptions. While it probably works well for modest pre-test differences, there is no guarantee that it corrects properly for gross differences between treatment and control groups, say those approaching or exceeding one standard deviation. Since researchers are reluctant to compare the growth patterns of white and black students precisely because their differences approach this magnitude, I question whether it makes sense to compare two groups of black students who exhibit similar differences.

Second, the guidelines do not require equivalent pre- and post-tests, but only that the content is similar and that the same test is used for both treatment and control groups. For example, SRA reading might be used as the pre-test and Iowa reading as the post-test. Although one can convert each test score to a standardized score, using that test's standard deviation, this converted mean still reflects test content, thereby preventing us from establishing that the treated group actually changed on the criterion in question. Moreover, if this issue is combined with substantial pre-test differences, it is quite possible that spurious effects can arise (e.g., high-achieving black students can show greater relative gain from the CTBS at time 1 to the Stanford at time 2 than low-achieving black students, and more than high-achievers would show from CTBS at time 1 to CTBS at time 2.

Fortunately, only one study (Rentsch, 1967) embodies both features, and, accordingly, I have excluded it from the review in the next section. I have also excluded the Thompson and Smidchens (1979) study on two grounds: its segregated control group averages only 42 percent black, which means it is not segregated by the 50 percent criterion, and no pre- or post-standard deviations are available for the purpose of computing a standardized effect estimate. A sensitivity analysis is shown in the discussion section to test the impact of these exclusions on my results.

Analysis Procedures

The fact that pre-test differences have a high positive correlation with post-test differences in the studies being reviewed makes it imperative to adjust post-test scores for pre-test differences. If this is not done, then desegregation effect estimates will be biased by pre-existing differences between segregated and desegregated students.

In general, I have followed the procedures outlined by Wortman (1982), with several refinements which are described here. Ideally, what one would like to have is a population standard deviation for each grade and test, so that truly standardized means could be calculated independent of sample variations. Unfortunately, this information is not readily available, and it is not available at all if one wishes to use estimates for black populations alone. Therefore, sample estimates of standard deviations must be used for calculating adjusted effect estimates.

My procedure differs from Wortman's only in the fact I pooled standard deviations wherever possible to improve the reliability of the standard deviation estimate. If the data shows an apparent fan-spread effect, indicated by higher post-test standard deviations than pre-test standard deviations, then standardized effects were computed separately for time 1 and time 2 means using pooled standard deviations for each time. If no fan spread was apparent, then all standard deviations were pooled for the estimate.

Moreover, I made estimates even where some or all sample standard deviations were missing. If only pre- or post- test standard deviations were available, then they were pooled for the population estimate. In a couple of instances I used standard deviation estimates from other studies in our NIE set, providing they were based on the same test. The advantage of this approach is that a greater number of adjusted effect estimates are available than in Wortman's approach. This analysis feature is fairly critical, since many otherwise excellent studies in our set have all of the design requirements and the pre- and post- test means, but lack only standard deviation estimates (sometimes from only one time period). It seems improper to exclude such studies from effect size means when other standard deviation information can be used to provide reasonable approximations.

Other less important analytic issues will be raised in the study-by-study discussion, to which we now turn.

REVIEW OF THE STUDIES

A summary of desegregation effects on black achievement from each of the 17 studies reviewed is tabulated in Table 1. More detailed information, including pre-test means, gain scores, and pooled standard deviations are shown in an appendix table, along with Wortman's effect estimates (which are very close to mine in most instances where he computes them). Table 1 also shows the results of significance testing carried out by each study's author, denoted by an asterisk next to the effect estimate if it exceeds the .05 level.

Anderson

The first study in the group, a voluntary transfer plan in Nashville, shows the largest effect sizes of the studies reviewed, for both math and reading. It is not only statistically significant (by the author's test), but educationally large as well, with reading gains nearing 1 standard deviation. Note that the study has converted test scores into T-scores relative to each grade level, so that decreases in the means are not inconsistant with increases in raw score means. Also, given this type of standardization, fan spread cannot be detected and so all sample standard deviations were pooled for the estimate. Since the two groups were equal on pre-test means, fan spread should not be a problem in any event.

Beker

This study evaluates a voluntary transfer plan in the North. Our analysis differs somewhat from Wortman (other than using pooled standard deviations). Wortman used a

TABLE 1

SUMMARY OF THE EFFECTS OF DESEGREGATION ON BLACK ACHIEVEMENT

Study Author	Grade Levels Tested Pre - Post[a]	Desegregation Effect Sizes[b] Reading	Math
Anderson	2S - 4S	+.89*	+.54*
Baker	2F - 2S	+.34	-.28
	3F - 3S	+.17	-.04
Bowman	3F - 5S[c]	+.03*	-.05
	3F - 5S[c]	-.55	-.37
Carrigan	KS - 1S	-.55	--
	1S - 2S	+.13	--
	2S - 3S	-.19	--
	3S - 4S	+.21	--
	4S - 5S	+.10	--
	5S - 6S	-.11	--
Clark	6F - 6S	-.01	-.12
Evans	4F - 4S	-.03	-.12
	5F - 5S	+.06*	+.26*
Iwanicki	2S - 3S	.00	--
	4S - 5S	.00	--
	6S - 7S	.00	--
Klein	10F - 10S	.00	-.08
Laird & Weeks	1S - 4F	+.54*	.00
	3F - 5F	+.24*	-.18
	4F - 6F	+.19	.00
Savage	9 - 11	+.15	-.08
Sheehan	4F - 5S	-.16*	-.21*
Slone	4S - 5S	+.27	+.47*
Smith	6S - 9S	-.06	+.13
Syracuse	4F - 4S	+.75*	--
	3F - 4S	.00	--
Van Every	4F - 6S	-.46	+.51
Walberg	3,4F - 3,4S	-.02	--
	5,6F - 5,6S	-.21	--
	7,9F - 7,9S	+.08	--
	10,12F-10,12S	-.25	--
Zdep	2F - 2S	+.53	-.17

* Significant at .05 level or better by author's test
a S denotes spring, F denotes fall
b In standard deviation units
c First entry uses regular segregated control group; second entry uses a segregated control group with an enriched program.

control group of black students who were accepted for the voluntary transfer plan but who ultimately turned it down. There was another potential control group of students who were accepted, but could not be accommodated in the transfer program due to lack of space. Since this group did not differ to any significant degree from the "refuser" group, I pooled the two groups to improve N's and standard deviation reliabilities. Compared to Wortman, this procedure yielded higher effects for reading but lower effects for math. The author did not compute a formal test so far as I can discern, but his discussion implied significant positive effects for 3rd grade reading, significant negative effects for 2nd grade math, and no other significant effects.

Bowman

The Bowman study is the only one I have included which uses different pre- and post-tests (N.Y. State and Iowa, respectively). One reason I included it was the fact that the pre-test showed only modest differences between the desegregated and the control groups (about ½ standard deviation), and also because it has a second and novel control group: black students remaining in a segregated school and classroom but with an enriched educational program. Interestingly, while there are no large effects of desegregation compared to the regular controls (although the author reports a significant t-test for reading), there is a very large effect (non-significant according to the author) showing that segregated enriched students gained more than desegregated students. (In the Appendix all means are divided by their respective standard deviations, and therefore appear in standardized form). Sensitivity analysis shown later evaluate the effect of including or excluding the segregated-enriched control group.

Carrigan

The Carrigan study evaluates a mandatory "one-way" busing program, arising from the closure of a predominately black school. One might object to the control group here, because it was just at 50 percent black. Nonetheless, it was in an area undergoing transition and does just barely meet the definition being used here.

Pre-test means are not shown in the Appendix, since Carrigan did not tabulate them for subjects in the study for both the pre- and post-test (there were some dropouts and missing data) Given the small N's such inconsistencies might bias the standard deviation estimates, so I simply pooled all standard deviations for a single estimate, which can then be divided into the gain score for the effect size. Wortman apparently used the existing pre- and post-standard deviations (with inconsistant N's), thereby accounting for the variations with my estimates. However, the estimates averaged across all grades are very close.

Clark

Clark evaluated a voluntary transfer program in North Carolina. This is the first study in the NIE set where all design criteria are met except pre- and post- standard deviations. Presumably because of missing standard deviations, Wortman analysed the SCAT verbal test; although even here only a single standard deviation is available. I have chosen the STEP reading test, although the results are similar to those for the SCAT. For a pooled standard deviation I have used the estimate from Savage (see below) whose standard deviation averaged 14 at the 9th grade level. According to STEP norm tables, the 6th grade standard deviation should be about 1 point lower than the 9th, but I have used 14 from Savage as a conservative estimate. Given the small change, a standard deviation in the 13 to 15 range will not alter the effect estimate. I also used 14.0 as the standard deviation for the SCAT quantitative test, although this is probably conservatively high (thereby producing a smaller negative effect). Fan spread should not be a problem here, since pre-test means are virtually identical for the two groups.

Evans

This study evaluates a comprehensive, two-way mandatory program in Ft.Worth, one of only two such programs in the NIE set. Again, all design requirements were met except for pre- and post- standard deviations, so we used those from Sheehan, who assessed black outcomes at the same grades in the sister city of Dallas (using the same test). I interpolated for an estimate of 4th grade Spring and 5th grade Fall. It should be noted that all standard deviation values here are lower than those shown for national norms.

Iwanicki and Gable

This study is the only one of several evaluating Project Concern, a voluntary program in New Haven, Conn. that qualified under the panel's guidelines. Unfortunately, this study focuses on the second year of desegregation, so this factor should be taken into account when interpreting the results. Considering the similarity of the pre-treatment means at each grade level, however, (which reflect the end of the first year of desegregation), and the fact that the control group was drawn randomly from a group meeting Project Concern's requirements, including agreeing to participate when an opening occurs, it appears there were no first-year effects either.

The study does not include standard deviations, but assuming that black students gain anywhere from ½ to 1 standard deviation in one year (more in earlier years), which is the pattern in our data, then the standard deviations are probably in the 10 - 15 range. This assumption is consistent with white student means reported by Iwanicki which are anywhere from 11 to 18 points higher than the black means. In any event, since the similarity of pre-test means diminishes the concern for fan spread, and since the gains are identical for grades 2 and 4, the effect size for those grades will be 0 regardless of the standard deviation estimate. For grade 6 we used a conservative effect estimate of 0, even though the effect would be negative if we had a specific standard deviation estimate.

Klein

This study of voluntary transfers in the South is one of only two studies in our set at the high school level. Two control groups were available, one randomly selected from all-black high schools and one matched on I.Q.. The latter group was selected, due to clear selection effects when transferees were compared to the randomly selected controls. We still have a pre-test difference of 7 points, but it would be 11 points if the random group was used. Only a single standard deviation is available from an analysis of variance table, so the possibility of fan spread cannot be taken into account. However, since the control group has a lower pre-test mean and since each group gained the same amount, any fan spread effect should change our 0 effect into a negative effect, thereby making 0 a conservative estimate.

Laird and Weeks

This Philadelphia study evaluates a voluntary program brought on by overcrowding in a black school. Students were bused to one of two white schools, Day and McCloskey. The black students bused to Day were highly biased compared to control students, with both IQ and pre-test means averaging at or near 1 standard deviation above the controls in grades 4 and 5 (in fact, their IQ's equalled white means in the receiving schools). Therefore the McCloskey students were selected for analysis. Since post-test standard deviations differed considerably from pre-test standard deviations, time-specific effect estimates were derived.

The effects in this study are quite large and significantly positive for reading at grades 4 and 5, but negligible and non-significant for math at all grade levels. The authors used matched samples for their significance tests.

Rentsch

The results from this two year evaluation of volunteer busing program in Rochester (grades 3,4, and 5) are excluded from Table 1 on methodological grounds. First, the pre-test and post-test were different tests, and the author did not make it clear which tests were used and when they were administered. Second, pre-test differences between the desegregated and segregated control groups neared or exceeded 1 standard deviation. Most devastating of all, information received after the panel had selected this study revealed that white students were included in the study, and the selection method used for the bused students makes it highly likely that the desegregated group had two to three times as many white students as the control group. This possibility could explain why the desegregated group had such higher pre-test means.

The average reading effect for the three grades in the Rentsch study is +.50, while the average math effect is -.11. Sensitivity analysis will show the effect of including or excluding this study on my overall conclusions.

Savage

This evaluation of a Richmond, Virginia voluntary evaluation plan is the only other study in our set to investigate the high school level. Three of the four standard deviation for reading were about equal and similar to published norms, but a fourth was 2½ times larger (post-test for controls) reflected a possible computational or typing error. These three standard deviations were pooled for reading; pooling was done separately for pre- and post- standard deviations for math due to fan-spread indications.

Sheehan

This study of the Dallas plan may be especially significant because of its large N (nearly 2,000 students), a time span of two years, and being the only other evaluation of comprehensive two-way mandatory busing in this set. While the negative effect of desegregation is not large here, the size of the N renders it statistically significant -- the only such negative effect in the set.

Slone

An example of pairing is illustrated in this New York City evaluation, although it was implemented in only a few schools. The desegregation started in Fall, 1964, but the pre-test was given in Spring, 1965, so this study also represents a test of second year effects. On the other hand, Slone presents reading tests from Spring, Grade 3 (1964) showing that the desegregated and segregated groups started out with the same relative difference in reading achievement (25.5 months vs. 21.5 months) prior to desegregation. These pre-test differences of about ½ standard deviation would make a pre- and post- standard deviations desireable, but they are not available. Only a single pooled standard deviation is used for the effect estimate.

Smith

This Tulsa, Okla. study is the only one in the NIE set to study school desegregation due to residential patterns; it is also one of the longest-term studies. The desegregated schools have a higher proportion black than the other studies, averaging about 42 percent.

Syracuse

This study evaluated an "open enrollment" busing program in Syracuse, New York. Matched and unmatched controls were available; only the matched groups were used here. The control group for the 4th grade group was drawn from a different school than attended by the bused students originally. An comparable to the desegregated group. No difference between pre- and post standard deviations was found, so one pooled estimate was used. Although Van Every reports a non-significant post-mean difference, there appears to be a calculation error. Both the reading and math diffences appear to be statistically different.

Walberg

This study evaluates the Boston METCO program, a voluntary city-to-suburb busing plan like Project Concern. Grades 3 and 4 are combined, as are 4 and 5, and so on, due to small N's in the control subjects. No differences between pre- and post-standard deviations were observed, so over-all pooled estimates are used at each grade level. Math results are unreported here because of unreadable figures on xeroxed copy.

Zdep

The final study evaluates another voluntary metropolitan plan. The pre- and post-tests are from the same publisher, but the two different forms are not directly comparable and hence the raw score "gains" presented in Table 1 are presented only so the reader can derive post-treatment means. When converted to standardized "scale" scores from published norms, the bused group gained 4 more points on reading and lost 2 on math when compared to the control group (the national standard deviation of the scale scores is 10). Zdep found one of the largest effects on reading in the set, but the small N renders it statistically non-significant.

The Wortman Effects

The Wortman formula always computes effect estimates seperately for time 1 and time 2, and uses only the control group standard deviations. One can see from the Appendix that whenever identical groups and tests are being assessed, in most cases my estimate agrees closely with Wortman's. The main discrepancies arise in the Carrigan and Walberg studies, where absence of pre- and post-means on the same group of persons led me to use only the gain scores and a pooled standard deviation. Even for these studies the effect estimate averaged across all grade levels is very similar. The discrepancy in the Beker study arises because I combined two groups of segregated students for the control group: those who "refused" to join the busing program, the group used by Wortman, and those who accepted but could not be accommodated.

The important difference between the Wortman formula and the approach used here is the number of effect estimates obtained. By pooling standard deviations and by extimating standard deviations from other information, effect estimates are obtained for every study. Even though a precise standard deviation is not available, in many cases the treatment-control initial scores and gain scores are so similar that the effect will be near zero no matter what standard deviation is used. These near-zero effects can have a significant impact on overall effect estimate averages.

DISCUSSION

Although the number of studies in the set reviewed here is not large, the advantage of the panel's approach is that most studies exhibit above-average methodology, and most appear to be carefully conducted. Most important, each study meets reasonable standards for plausible causal inference: a pre-post design with a control group. What is lost in numbers, then, is gained in design quality, which is essential in arriving at a sound judgement about the impact of desegregation on black achievement.

The studies also exhibit a variety of desegregation settings and types, although they are weighted more towards voluntary programs than mandatory, a definite limitation for generalization. On the other hand, for this reason this set may provide a good test of the hypothesis, since it is probably the case that voluntary programs offer better opportunities for positive effects more support from the community, self-selection of families most desirous of the experience, and so forth.

The other major restriction on generalization is that the longest-term study here is only three years in duration, thereby complicating inference for desegregation experience spanning the whole school cycle. Given this panel's search, apparently there are no longer-term studies of adequate quality.

Taken as a whole, what do these studies tell us about desegregation and black achievement? There are several ways to approach an answer to this question.

First, we can consider the significant tests carried out by the author of each study. Of the 47 different grades and tests in these studies that were subjected to statistical analysis, only 11 were found significant at an acceptable level, and two of these were negative effects. We would add three more significant results out of 53 possible if the Rentsch study were to be added to the set. Thus the overwhelming majority of these studies, taken individually, found no significant effects of desegregation on black achievement.

The meta-analysis technique employed by the panel provides a second and more reliable method that goes beyond this simple counting exercise. We can arrive at an overall assessment of desegregation's impact by averaging the size of effects across all studies and grade levels. I adopted two alternative strategies in computing these overall averages. First, I computed the average of the effect estimates shown in Table 1, which reflects a group of studies that differs somewhat from the total group adopted by the panel. Second, for sensitivity purposes, I averaged effects for the original set of studies as selected by the panel. This second set of averages therefore includes results from the Rentsch study and the Thompson and Smidchens study and excludes the extra grades I analysed from the Bowman and Syracuse studies.

The average effect sizes are shown in Table 2. For the set of studies I selected, the average effect is .06 of a standard deviation for reading and .01 for math. Neither of these two average effect sizes are significantly different from 0 by statistical test. When we consider those studies

TABLE 2

THE AVERAGE EFFECT OF DESEGREGATION ON BLACK ACHIEVEMENT

	Average Effect Size[a]	
Study Grouping	Reading	Math
Table 1 Studies (N)[b]	.06 (33)	.01 (18)
Original Panel Studies (N)	.11 (35)	.00 (22)

a In fractions of standard deviation. One-tenth of the black student standard deviation (.10) is equivalent to about one month of educational growth as measured by most standardized tests.

b Number of grade levels for which the average is computed.

as originally adopted by the panel, the effect for reading rises to .11 and the math effect falls to 0. The reading effect is still not significantly different from 0. The average reading effect size of .11 for the panel's original studies is somewhat smaller than Wortman's average effect, primarily because of his decision not to calculate effect estimates for a number of studies with effects near 0 (due to incomplete standard deviation information).

For the sake of discussion, let us assume that the more liberal effect estimate of .11 for reading held up across a larger number of studies, so that it would be statistically significant. We must still decide whether a reading effect of this size would be educationally significant.

First, we must keep in mind that the unit of measurement here is variation in black scores, which is known to be smaller than that for black and white students combined, or for national norm data, perhaps on the order of two-thirds or three-fourths. Therefore, even if one found an effect of .11 in a larger group of studies, the effect in terms of national norms is still less than .10 or less than one month of a school year. Since the achievement differencial between black and white students averages between 1 and 1.5 standard deviations, an average effect of .11 for black reading achievement means that desegregation alone could close the gap by less than 10 percent.

Second, such an effect might be educationally significant if it was cumulative over time; that is, if a black child gained

.11 or one month of a school year for each year the child was in a desegregated school. Is there any evidence for such a possibility in this group of studies? This possibility can be tested to some extent by dividing up studies according to duration and computing average effects for one-year studies, two-year studies, and three-year studies. I have carried out this analysis for reading scores using the panel's original 35 grade levels. If desegregation effects are cumulative, one should see increasing effects sizes as the duration of desegregation increases.

The results for reading are summarized in Table 3. The average effect is +.04 for one-year studies, +.37 for two-year studies, and -.16 for three-year studies. While the two-year studies do have larger effects on the average than one-year studies, the three-year studies show an average negative effect (due largely to the Van Every study). Therefore, there is no evidence from these studies -- the best available -- that there is any cumulative effect of desegregation. This conclusion must be qualified, of course, by the fact of the relatively small number of cases for any given duration period.

What about the grade at which children are desegregated? When we compute average effects by grade level, the studies here reveal average effects of -.55 for desegregation begun at grade one (one study), .35 for grade 2, and inconsistant effects near zero for other grades. This set of high-quality studies does not support Crain and Mahard's finding of large effects for

TABLE 3

THE EFFECT OF DESEGREGATION ON BLACK READING ACHIEVEMENT, BY YEARS OF SEGREGATION[a]

Length	Average Reading Effect Size	
One year	+.04	(N=23)
Two years[b]	+.37	(N=9)
Three years[c]	-.16	(N=3)

a Using only the original panel studies, including Rentsch and Thompson & Smidchens.

b Anderson, Laird & Weeks, Rentsch, Savage and Sheehan.

c Bowman, Smith and Van Every.

documented in these studies. We have no way to investigate what these factors might be, but one hypothesis is that they are due to unique educational programs available in those few schools. Indeed, given the much larger effects demonstrated in many purely academic interventions (see Walberg's paper in this volume for a discussion of some of these interventions), this hypothesis may be the only reasonable explanation for the considerable variation observed in the panel's selected studies.

IMPLICATIONS FOR POLICY

Although the findings of each paper in this volume differ to some extent, the range of difference is small in comparison with previous debates on this issue. With the exception of Crain, all panelists find no effects for math achievement, and find that reading effects are positive but quite small and not educationally significant in all but a few studies. Perhaps a majority of the panel also agrees that the average reading effects are considerably smaller than what might be expected from special educational interventions.

What, then, should the policy directions be from this consensus of experts? It seems to me there are four audiences whose future actions might be influenced by these results.

The community of educational researchers might justifiably decide that enough research has been done on the issue of desegregation and achievement, and that their energies and resources should be devoted to more fertile pastures. There will be some, of course, who will find sufficient flaws in all 19 of these "best" studies to recommend one more large-scale, well-funded study to provide a definitive answer. I would not quarrel with such a study, but at this point the probability of a negative or indeterminate answer (given current knowledge) is high, thereby making its cost hard to justify.

For educational policy makers, I think these results offer an excellent opportunity to reconsider priorities for programs designed to enhance minority student achievement. Desegregation is simply not a cost-effective technique to accomplish this goal. However desireable racial balance may be for other purposes, it is not going to reduce the achievement differential between white and black students. It is time to solve educational problems with educational solutions, and many promising directions are documented in the Walberg paper.

The courts and civil rights activists should also take note of these findings. The studies reviewed here tell us nothing about whether segregation caused the black-white achievement gap, but they do tell us that desegregation by itself will not close it to any important degree. There is controversy about the role played by achievement issues in the original Brown decision, but there is no question that many lower courts have been influenced by achievement results when fashioning desegregation remedies. One hopes that the results here will relieve judges of the misconception that they are benefiting the academic progress of minority students by ordering desegregation plans.

Finally, these findings may offer relief to many black parents who have willingly endured the hardships of cross-town school transfers because of the mistaken belief that their children will benefit academically. Many will continue to endorse such transfers for other reasons, but many others may well be happy to discover that their child can get just as good an education in a neighborhood school close to home.

This does not mean we should abandon desegregation: it remains a goal all panel members share. I think it does raise serious questions about compulsory desegregation methods such as mandatory busing. There is little justification for forcing parents and children into expensive, time-consuming cross-town bus rides when there is no educational advantage. For those of us who want to pursue the goal of integrated education, we should support comprehensive voluntary transfer programs, on a metropolitan basis where necessary. It should be made clear to all participants, however, that simply changing to schools that are more racially balanced than one's neighborhood school is no guarantee of a superior education. Indeed, they may be giving up possible advantages of special programs in their own school -- programs designed specifically to enhance education and proven to work.

THE EFFECT OF DESEGREGATION ON BLACK READING AND MATH ACHIEVEMENT

Study and Grade/Year	Test and (N_D/N_S)	Desegregated Pre $\bar{X}$	Desegregated $Gain_D$	Segregated Pre $\bar{X}$	Segregated $Gain_S$	$Gain_D$-$Gain_S$	Pooled sd (T_1/T_2)	Effect	Wortman Effect	Author Test
Anderson	Metro (T-scores)									
2/60 - 4S/63	(34/34)	44.3	2.3	46.4	-4.8	+7.1	8.0	+.89	+.95	+
MATH:	(34/34)	44.6	3.6	43.8	-1.3	+4.9	9.0	+.54	+.53	+
Beker	Stanford (GE months for paragraph meaning)									
2F/64 - 2S/65	(25/32)	15.9	6.7	16.3	5.2	+1.5	2.3/6.7	+.34	+.23	?
3F/64 - 3S/65	(11/28)	24.2	8.5	20.0	5.5	+3.0	6.6/8.9	+.17	-.04	?
MATH:	(25/32)	15.6	4.7	16.7	7.1	-2.4	4.3/6.6	-.28	-.02	?
(Concepts)	(11/28)	20.6	7.6	20.3	7.9	-0.3	6.9/9.3	-.04	+.59	?
Bowman	Iowa (Pre-test is NY State; scores here are standardized by test sd's)									
3F/67 - 5S/70	(12/36)	2.80	-.06	2.33	-.09	+.03	4.7/12.0	+.03	+.02	+
	(" /21)	(Seg. Enriched)		2.24	+.61	-.55		-.55	--	0
MATH:	(12/38)	2.16	+.14	2.05	+.19	-.05	2.7/7.0	-.05	-.06	0
	(" /21)			1.95	+.51	-.37		-.37	--	0
Carrigan	California (Age-equivalent)									
KS/65 - 1S/66	(17/23)		7.1		10.0	-2.9	5.3	-.55	-.41	0
1S/65 - 2S/66	(16/21)		7.5		6.7	+0.8	6.3	+.13	-.02	0
2S/65 - 3S/66	(25/23)		6.6		8.3	-1.7	9.0	-.19	+.30	0
3S/65 - 4S/66	(11/23)		11.6		9.2	+2.4	11.2	+.21	-.13	0
4S/65 - 5S/66	(13/24)		9.1		7.3	+1.8	17.4	+.10	+.33	0
5S/65 - 6S/66	(13/21)		3.3		5.1	-1.8	16.9	-.11	-.31	0
Clark	STEP (Coverted scores)									
6F/69 - 6S/70	(108/88)	250	4.9	248	5.1	-0.2	14.0*	-.01	--	0
MATH: (SCAT)	(108/88)	254	5.5	254	7.2	-1.7	14.0*	-.12	--	0
Evans	Iowa (GE months)									
4F/71 - 4S/72	(393/180)	32.0	3.0	29.0	3.0	0.0	10.0/11.6**	-.03	--	0
5F/71 - 5S/72	(381/181)	39.0	2.0	37.0	1.0	+1.0	11.6/13.2**	+.06	--	+
MATH:	(392/179)	33.0	4.0	32.0	5.0	-1.0	8.3/9.8**	-.12	--	0
	(386/181)	40.0	5.0	39.0	2.0	+3.0	9.8/11.3**	+.26	--	+

-38- APPENDIX

Study and Grade/Year	Test and (N_D/N_S)	Desegregated Pre $\bar{X}$	$Gain_D$	Segregated Pre $\bar{X}$	$Gain_S$	$Gain_D$ - $Gain_S$	Pooled sd (T_1/T_2)	Effect	Wortman Effect	Author Test
Iwanicki	Woodcock									
2S/76 - 3S/77	(64/50)	102	13	100	13	0.0	?	.00	--	0
4S/76 - 5S/77	(66/48)	125	5	124	5	0.0	?	.00	--	0
6S/76 - 7S/77	(70/65)	136	2	134	5	-3.0	?	.00	--	0
Klein	Cooperative									
10F/65-10S/66	(38/38)	104	13	97	13	0.0	31.6	.00	--	0
MATH: (Z-scores)	(38/38)	.23	.03	-.16	.11	-0.08	1.0	-.08	--	0
Laird & Weeks	Philadelphia Achievement									
1S/63 - 4F/65	(20/140)	3.7	5.1	4.2	4.0	+1.1	1.7/2.3	+.54	--	+
3F/63 - 5F/65	(13/140)	7.2	4.2	6.7	2.2	+2.0	1.4/2.5	+.24	--	+
4F/63 - 6F/65	(10/147)	8.4	4.1	9.1	3.7	+0.4	2.2/2.6	+.19	--	0
MATH:	(19/138)	4.9	2.3	5.6	2.9	-0.6	1.6/3.0	.00	--	0
	(16/139)	6.6	2.6	6.8	3.4	-0.8	2.0/2.9	-.18	--	0
	(14/167)	7.7	4.3	8.5	4.3	0.0	2.9/2.8	.00	--	0
Savage	STEP (Converted scores)									
9/68 - 11/70	(42/42)	269	10.6	271	8.5	+2.1	14.2	+.15	+.14	0
MATH:	(42/42)	256	3.6	253	3.8	-0.2	11.5/16.0	-.08	-.05	0
Sheehan	Iowa (GE months)									
4F/76 - 5S/78	(810/1115)	27.6	9.2	29.0	11.8	-2.6	10.0/13.2	-.16	-.16	-
MATH:	(810/1115)	28.3	8.2	29.2	10.3	-2.1	8.3/11.3	-.21	-.16	-
Slone	Metro (GE months)									
4S/65 - 5S/66	(86)	40.2	11.0	34.9	8.8	+2.2	8.1	+.27	--	0
MATH:	(98)	38.1	5.1	36.7	2.1	+3.0	6.4	+.47	--	+
Syracuse	Stanford (GE months)									
4F/65 - 4S/66	(24/24)	34.5	9.2	34.3	4.0	+5.2	7.2	+.75	--	+
3F/64 - 4S/66	(12/12)		11.4		11.4	0.0	8 to 9	.00	--	0

Study and Grade/Year	Test and (N_D/N_S)	Desegregated Pre $\bar{X}$	$Gain_D$	Segregated Pre $\bar{X}$	$Gain_S$	$Gain_D$ - $Gain_S$	Pooled sd (T_1/T_2)	Effect	Wortman Effect	Author Test
Smith	Stanford (Raw score for paragraph meaning)									
6S/65 - 9S/68	(124/150)	16.8	18.5	18.1	19.7	-1.2	8.8/12.0	-.06	-.05	0
MATH: (Comput. raw)	(124/150)	10.5	12.3	9.3	10.5	+1.8	4.1/7.2	+.13	+.10	0
Van Every	SRA (GE months)									
4F/66 - 6S/69	(20/21)	31.6	11.5	29.4	16.2	-4.7	10.3	-.46	-.44	0
MATH:	(20/21)	29.6	19.0	30.8	15.2	+3.8	7.4	+.51	+.53	0
Walberg	Metro (Raw)									
34F/68-34/69	(90/17)		1.8		2.0	-0.2	7.9	-.02	+.11	0
56F/68-56/69	(61/29)		3.6		5.0	-1.4	6.8	-.21	-.24	0
79F/68-79/69	(124/25)		2.1		1.5	+0.6	7.8	+.08	+.21	0
HSF/68-HS/69	(72/14)		1.7		3.2	-1.5	6.0	-.25	-.01	0
MATH:										
Zdep	Coop. Primary (Raw scores--pre is 12A, post is 23A)									
2F/68 - 6S/69	(12/15)	14.5	8.4	16.0	4.5	+3.9	6.9/7.8	+.53	+.65	0
MATH:	(12/15)	26.3	-1.9	26.3	-1.0	-0.9	6.8/5.4	-.17	-.15	0

*Estimated from Savage **Estimated from Sheehan

BIBLIOGRAPHY

Anderson, Louis V. The effect of desegregation on the achievement and personality pattern of Negro Children. Ph.D. dissertation, George Peabody College for Teachers (University Microfilms No. 66-11237).

Armor, David J. The Evidence on Busing. Public Interest 28, 90-126, 1972.

Armor, David J. The Double Double Standard: A Reply, Public Interest, 30, Winter, 1973.

Beker, Jerome A study of integration in racially imbalanced urban public school, Syracuse, New York: Syracuse University Youth Development Center, Final Report, May 1967.

Bowman, Orrin H. Scholastic development of disadvantaged negro pupils: A study of pupils in selected segregated and desegregated elementary classrooms. Unpublished doctoral dissertation, University of New York at Buffalo 1973 (Microfilm No. 73-19176).

Bradley, L.A. & Bradley, G.W. The academic achievement of black students in desegregated schools: A critical review. Review of Educational Research, 1977, 47, 399-449.

Carrigan, Patricia A. School desegregation via compulsory pupil transfer: Early effects on elementary school children. Ann Arbor, Michigan: Ann Arbor Public Schools, 1969.

Clark, El Nadel Analysis of the difference between pre- and posttest scores (change scores) on measures of self-concept, academic aptitudes, and reading achievement earned by sixth grade students attending segregated and desegregated schools. Unpublished doctoral dissertation, Duke University, 1971.

Crain, R.L. & Mahard, R.E. Desegregation plans that raise black achievement: A review of the research. Santa Monica, CA: The Rand Corporation (N-1844-NIE), June 1982.

Evans, Charles L. Integration evaluation: Desegregation study II -- academic effects on bused black and receiving white students, 1972-73. Forth Worth, Texas: Fort Worth Independent School District, 1973 (ERIC No. ED 094 087).

Glass, G.V. Primary, secondary and meta-analysis of research. Educational Researcher, 1976. 5, 3-8.

Iwanicki, E.F. & Gable, R.K. A quasi-experimental evaluation of the effects of a voluntary urban/suburban busing program on student achievement. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, Canada, March 1978.

Klein, Robert S. A comparative study of the academic achievement of negro tenth grade high school students attending segregated and recently integrated schools in a metropolitan area in the south. Unpublished doctoral dissertation, University of South Carolina, 1967.

Krol, R.A. A meta-analysis of comparative research on the effects of desegregation on academic achievement. Unpublished doctoral dissertation, 1978, Western Michigan University: University microfilms (No 79-07962), 1979.

Laird, M.A. & Weeks, G. The effect of busing on achievement in reading and arithmetic in three Philadelphia Schools, Philadelphia, Pennsylvania: The School District of Philadelphia, Division of Research, 1966.

Pettigrew, T.F. Busing: A review of the Evidence, Public Interest, 30, Winter 1973.

Rentsch, George J. Open-enrollment: An appraisal. Unpublished doctoral dissertation, State University of New York, Buffalo, 1967.

Savage, L.W. Academic achievement of black students transferring from a segregated junior high school to an integrated high school. Unpublished masters thesis, Virginia State College, 1971.

Sheehan, Daniel S. Black achievement in a desegregated school district: Journal of Social Psychology, 1979, 107, 185-192.

Slone, Irene W. The effects of one school pairing on pupil achievement, anxieties and attitudes. Unpublished doctoral dissertation, New York University 1968.

Smith, Lee R. A comparative study of the achievement of negro students attending segregated junior high schools and negro students attending desegregated junior high schools in the city of Tulsa. Unpublished doctoral dissertation, University of Tulsa, 1971.

St. John, N.H. School desegregation: Outcomes for children. New York: John Wiley & Sons, 1975.

Syracuse City School District. Study of the effect of integration -- Washington Irving and Host pupils. Hearing held in Rochester, New York, September 16-17, U.S. Commission on Civil Rights 1966, pp 323-236.

Thompson, E.W. & Smidchens, U. Longitudinal effects of school racial/ethnic composition upon student achievement. Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, California, April 1979).

Van Every, D.F. Effect of desegregation on public school groups of sixth graders in terms of achievement levels and attitudes toward school. Doctoral dissertation, Wayne State University, 1969. Dissertation Abstracts International 1969, (University Microfilms No. 70-19074).

Walberg, Herbert J. An evaluation of an urban-suburban school busing program: Student achievement and perception of class learning enviroments. Paper presented at the annual meeting of the American Educational Research Association, New York: 1971.

Walberg, Herbert J. Desegregation and Educational Productivity National Insitute of Education, 1983.

Weinberg, M. Desegregation Research: An Appraisal. Bloomington, Ind. Phi Delta Kappa, 1970.

Weinberg, M. Minority Students: A research appraisal. Washington D.C.: U.S. DHEW, National Institute of Education, 1977.

Wortman, Paul M. School Desegregation and Black Achievement: An Integrative Review. University of Michigan, 1983.

Zdep, Stanley M. Educating disadvantaged urban children in suburban schools: An evaluation. Journal of Applied Social Psychology, 1971, 1, (ERIC, No. ED 053 186 TM 00716).

grade 1 (and kindergarten) but no effects for grade 2 and higher grades.

Finally, it is noted that there are several studies with very sizable reading effects: Anderson, Syracuse, Zdep, one grade from Laird and Weeks, and two grades from Rentsch. Without these six grades (out of 35 in the set), the reading effect would be near 0. Therefore, even the overall average reading effect of .11 is not a consistant effect of desegregation. It would be more acurate to summarize our studies by saying there are six grades with substantial reading effects ranging from .5 to .8 and 29 grades with much smaller reading effects that average out to about 0.

No matter how one summaries these desegregation effects, the conclusion is inescapable: the very best studies available demonstrate no significant and consistant effects of desegregation on black achievement. There is virtually no effect whatsoever for math achievement, and for reading achievement the very best that can be said is that only a handful of grade levels from the 19 best available studies show substantial positive effects, while the large majority of grade levels show small and inconsistant effects that average out to about 0.

The fact that only a small fraction of these studies show substantial effects, even though all grade levels were desegregated, suggests strongly that factors other than desegregation are the real causes of the large achievement gains