An Evaluation of

AMERICANS ALL

A National Multicultural Education Program

Results of an Evaluation of Program Effects

in Washington, D.C.

Mark F. Testa, Ph.D.

Associate Professor

School of Social Service Administration

The University of Chicago

February 1992

The changing demography of American classrooms is prompting a nationwide movement to restructure school syllabi to reflect society's growing ethnic diversity. The likely dimensions of this change are already discernible in the ethnic profiles of the five states with the largest populations under 18 years of age- California, Florida, Illinois, New York, and Texas. Together they account for 35 percent of the nation's child population. Presently, 43 percent of the children of these states do not have European ethnic origins. Rather, they are: Hispanic (24%), African (14%), and Asian (5%). By the year 2000, children of these diverse ethnic origins will become the majority of students in these five states. This pattern could well become characteristic of the entire nation toward the middle of the twenty-first century if current demographic trends continue.

To prepare students to function in a multicultural society, educators have begun to diversify school syllabi to acquaint students with the history of minority and diverse ethnic and racial groups whose experiences have hitherto been neglected or omitted from classroom curricula. Although many people welcome the diversification of curricula as enriching students' cultural identities, self-concepts, and multicultural awareness, others warn that the assumptions and goals of the current reforms risk undermining some essential foundations of American solidarity. Certain commentators caution that by schools' celebrating and strengthening children's sense of ethnic origins and identities, society runs the danger of encouraging social divisiveness at the expense of national unity (Ravitch 1990, Schelsinger 1991).

Progress toward settling such issues has been hampered by a lack of empirical studies of programs of multicultural education. Although many of the arguments in favor of a more diversified curriculum have been around for several decades, some of its key claims, such as improv-

ing student self-esteem and strengthening scholastic interest, still must await research verification. Similarly, the concern that multicultural education undermines national identification and promotes separatism remains only speculative at best.

To advance our understanding of these issues, it is important to conduct careful evaluations of programs of multicultural education and assess their effects on students and consequences for their education. This article reports findings from a study of the implementation of one such multicultural education program, Americans All, which was field-tested in the Washington D.C. public school system in 1991.

Background

Sleeter and Grant (1987) identified five distinct approaches to multicultural education which emerge from the literature: (1) "to assimilate minority students into the cultural mainstream and existing social structures by offering transitional bridges within the existing school program"; (2) "to help students of different backgrounds get along better and appreciate each other"; (3) "to foster cultural pluralism by teaching courses about the experiences, contributions, and concerns of distinct ethnic, gender, and social class groups"; (4) "to promote cultural pluralism and social equality by reforming the school program for all students to make it reflect diversity"; and (5) "to prepare students to challenge social structural inequality and to promote cultural diversity."

Americans All reflects each of the above approaches in varying degrees, but its primary emphasis is on fostering cultural pluralism by teaching the history of both voluntary immigration to the United States and the involuntary incorporation of various ethnic and racial groups through colonization, enslavement, or territorial annexation. The program is intended to augment the standard social studies curricula for kindergarten through the twelveth grade. It utilizes materials

from the Images of Ellis Island curriculum which commemorates the American immigrant experience and the historic importance of Ellis and Angel Islands. It supplements these materials with essays on the particular experiences of African Americans, Asian Americans, European Americans, Mexican Americans, Native Americans, and Puerto Ricans. Teachers are trained in a two-day workshop in the use of these products and in methods of appeciative learning to promote four general goals: (1) enriching children's cultural identities, (2) enlarging their multicultural awareness, (3) enhancing self-esteem, and (4) fostering critical thinking skills.

A core value of Americans All and other multicultural programs is cultural pluralism. The following two statements are indicative of this orientation. The first comes from the Introduction to the Americans All's teachers guide and second from the National Coalition for Cultural Pluralism:

> The world in general and American society in particular are made up of many different groups of people with a variety of familial, ethnic and cultural backgrounds. Working cooperatively and effectively within communities is easier when people appreciate and value these multiple cultures and human experiences. Developing multicultural awareness increases one's sense of security or belonging, provides opportunities for expression and communication, and enhances one's sense of personal empowerment and freedom (Christopher and Sreb 1989, p.*i*-1).

> Cultural pluralism is a state of equal co-existence in a mutually supportive relationship within the boundaries or framework of one nation of people of diverse cultures, with significantly different patterns of belief, behavior, color, and in many cases with different languages. To achieve cultural pluralism, there must be unity within diversity. Each person must be aware of and secure in his own identity, and be willing to extend to others the same respect and rights that he expects to enjoy himself (quoted in Suzuki 1979, p. 45).

Implicit in the above statements are a set of propositions that provide the rationale for Americans All and similar multicultural programs: (1) increasing students' awareness and appreciation of their own *cultural identity* and origins, (2) fashioning a basis for increasing their *multi-*

cultural awareness and respect for different cultural identities and origins, which (3) help to promote mutual regard and positive *self-esteem*, and, (4) expand the opportunities for communication and development of *critical thinking* skills.

While there is scattered research around to support each of the above propositions to varying degrees, much skepticism about their validity still remains. For example, some studies show that multicultural programs can simultaneously improve ethnic self-awareness and decrease ethnocentric attitudes (Project REACH 1990). Still, it is feared that by celebrating and strengthening children's cultural identities and origins schools could just as likely reinforce ethnic, racial, and religious divisiveness as promote pluralistic appreciation (Schlesinger 1991). Similarly, some social pyschological theories imply that multicultural education can improve student self-esteem by neutralizing the impact of negative stereotypes (Ransey 1982, Crocker and Major 1989). But recent research shows that the self-esteem of minority children is no worse or even slightly better than the self-esteem of majority white children (Bachman and O'Malley 1984). This research calls into question both the need for and capacity of multicultural programs to raise student self-esteem. Likewise, there is no guarantee that multicultural education will automatically foster critical thinking skills. In fact, many are worried about the opposite occurring: the promotion of a cultural relativism that can lead to an uncritical acceptance of all aspects of every culture, even those that are exploitive or oppressive (Suzuki 1979, Bloom 1987).

Acknowledging the potential limitations or unintended consequences of multicultural education programs does not argue for their abandonment. What is required in light of the current controversies, rather, is careful research into the anticipated goals of multicultural programs and a rigorous accounting of their actual effects on children. Findings from such research should not

be seen as justifying or undermining ongoing reform efforts but rather as contributing to a deliberative process that informs policy makers, funders, and administrators about past accomplishments and future adjustments that need to be made in order to achieve desired goals.

In the next section, I shall discuss the methods used in this study to evaluate the implementation of the Americans All program in the Washington, D.C. school system. The original design called for the random assignment of classrooms within schools to the Americans All programs or to a one-year waiting list (control group). Two-rounds of identical instruments were administered to students in the program and control groups. All students were asked to complete a brief, self-report measure designed to assess self esteem in children and adolescents. In addition, students in grades eight and eleven were asked to complete a questionnaire designed to measure scholastic interest, educational aspirations, perception of school climate, attitudes toward cultural pluralism, and their estimation of the contributions of different ethnic groups to American history. The primary hypothesis of this study is that children in the Americans All program will exhibit significant differences relative to a control group in self-esteem measures, ratings of minority ethnic groups' contributions to American history, and acceptance of cultural pluralism.

Method

Subjects

The subjects of this study are students attending grades 3, 5, 8, and 11 in Washington, D.C. public schools. In 1991, the principals of 23 school volunteered their institution's participation in the research. A total of 39 program classrooms and 39 control classroom out of an original 102 were deemed eligible for random assignment to the program or to a waiting list for participation in the next academic year. Classrooms on the waiting list constituted the control group.

A total of 10 program classrooms (206 students) and 10 control classrooms (180 students) ultimately elected to participate in the research by completing pretest instruments. Unfortunately, attrition from the study was high. Posttest instruments were completed by 7 program classrooms but only 3 control classrooms. Refusals by some students to provide identifying information limited the matching of pretest and posttest instruments, which further reduced the final sample size to 81 students in the program group and 69 students in the control group.

Although the study was originally designed to achieve equivalence between the program and control groups through random assignment of classrooms, the heavy attrition of participants makes it untenable to assume that the groups remain statistically equivalent. Therefore, my analysis assumes a nonequivalent control group design (Cook and Campbell 1979) and employs multiple regression (analysis of covariance) to adjust for pre-existing differences. While this fall-back option complicates the conclusions one feels confident in drawing about the effectivenesss of the Americans All program, the fact that pretest and posttest observations are available for both program and control groups does permit some reasonable interpretations of results to be made.

Variables and Instruments

Self-Esteem (Grades 3 and 5): The Piers-Harris Children's Self-Concept Scale (Piers 1984) was used to assess children's self esteem. The Piers-Harris scale is an 80-statement self-report inventory designed for use with students in grades 4-12. Children respond "yes" or "no" to indicate whether each statement is self-descriptive. Total scores range from 0 to 80 with higher scores indicating greater self-esteem. In addition to the total score, the instrument yields six factor analytically derived subscales: behavior, intellectual and school status, physical appearance and attributes, anxiety, popularity, and happiness and satisfaction. The instrument is gener-

ally regarded as psychometrically sound for assessing children's self esteen (Chiu 1988). In some quarters, however, it is considered better suited for children than adolescents (Blascovich and Tomaka 1991). Testing forms were purchased from Western Psychological Services.

Self-Esteem (Grades 8 and 11): Under licensing agreement with Western Psychological Services, the Piers-Harris scale was also included in a special questionnaire I developed for Americans All's use with eighth and eleventh graders. Because of the uncertainty over Piers-Harris's suitability for adolescents, the questionnaire also included a variant of the widely used Rosenberg (1965) Self-Esteem Scale. The Rosenberg scale was originally a 10-item inventory designed to measure adolescents' global feelings of self-worth. The particular version reproduced in the Americans All questionnaire is a modified version of the original scale which the U.S Department of Education used in its 1988 National Education Longitudinal Study (NELS:88) of eighth graders. The NELS:88 version was chosen over the original scale to facilitate drawing comparisions between the Washington D.C. sample and a 1988 national sample of eighth graders. Although the Rosenberg scale is typically scored using a Likert-like response format (strongly agree, agree, disagree, strongly disagree), the dichotomized format of the Piers-Harris scale was used in the Americans All questionnaire. This change was made partly for convenience sake, but also was motivated by research that pointed to the possibility of systemmatic biases in black student responses to Likert-type questionnaire items, who tend to favor the extreme response categories independent of item content. Bachman and O'Malley (1984) reported that black students scored significantly higher than whites in self-esteem scores when a full four- or five-point response range was used, but that the racial difference disappeared when a truncated scoring method was used. I obtained similar test results for the NELS:88 sample using a dichotomized scoring method to control for racial differences in the use of extreme response categories. A dichotomized scoring method will also be used in this study.

Cultural Identity (Grades 8 and 11): The Americans All questionnaire for grades 8 and 11 asked students to report their *ethnicity* by circling the one category that best describes their ethnic or cultural group. The choices were: (1) Black or African American, (2) White or European American, (3) Mexican American, Mexican or Chicano, (4) Cuban American or Cuban, (5) Puerto Rican American or Puerto Rican, (6) Asian American, Asian or Pacific Islander, (7) American Indian or Native American, or (8) Alaskan Native. Students were also asked to characterize their *own ethnic group's achievements* by choosing between two things about their ethnic or cultural group that made them feel most proud. The choices were: (1) scientific contributions or athletic achievements, (2) acting and musical accomplishments or contributions to American history, (3) political leadership or religious leadership, and (4) money-making abilities or academic achievements. The responses were recoded into popular achievements (athletic, acting and musical, political, money-making) and other achievements (scientific, historical, religious, academic) and summed together to form a total score. The total score ranges from 0 to 7 with higher scores indicating higher pride in non-popular achievements. Lastly, students were asked to assess the contributions of various ethnic groups to the country's history using a four-item response format (very much, some, a little, very little). The response for the group corresponding to their self-reported ethnicity was used to measure students' assessments of their *own ethnic group's historical contributions* to the country. The scores range from 0 to 3 with higher scores indicating greater contributions.

Multicultural Awareness (Grades 8 and 11): Unlike self-esteem measures, it is difficult to find instruments that assess people's orientation toward the value of cultural pluralism. Adorno et al.'s (1950) Ethnocentrism (E) Scale and Bogardus's (1950) Social Distance Scale measure closely related constructs, but neither was considered suitable for today's junior and senior high school population. Therefore, I helped Americans All to design a set of items (see Figure 1) to

Figure 1. Feelings About Peoples Differences

How much do you agree with each of the following statements about people's differences?

Items	Strongly Agree	Agree	Disagree	Strongly Disagree
	(Circle One On Each Line)			
1. People of different color (whether they are black, white, brown or some other color) are basically the same.	1	2	3	4
2. People who live here but don't speak English aren't true Americans.	1	2	3	4
3 It would be better for everyone if people dated only people of their own race.	1	2	3	4
4.* People's differences in language, religion and customs are the real strengths of this country.	1	2	3	4
5. People who weren't born in the United States haven't contributed much to this country's history.	1	2	3	4
6. It would be better all around if people spoke the same language, practiced the same religion, and shared the same customs.	1	2	3	4

* Scoring reversed to indicate pluralistic appreciation.

measure students' feelings about people's differences. For the analysis, I constructed two additive scales from the items: degree of *pluralistic appreciation* (items 1, 4 and 6) and degree of *pluralistic acceptance* (items 2 and 5). The scores range from 0 to 15 and from 0 to 7, respectively, with higher scores indicating greater appeciation or acceptance. Item 3 on interracial dating was analyzed separately. As an additional indicator of multicultural awareness, I computed a measure of students' assessments of *other ethnic group's historical contributions* by adding together their responses on the contributions of different ethnic group's to the country's history. The scores range from 0 to 19 with higher scores indicating greater contributions.

Background and Intervening Variables (Grades 8 and 11): In studies where the control of variables cannot be adequately attained through randomization, statistical techniques, such as multiple regression or covariance analysis, can be useful devices for adjusting for pre-existing differences that make the groups nonequivalent in some important respects. For this purpose, I included a number of questions in the Americans All questionnaire to provide data on the background characteristics of students in the program and control groups. Most of these questions were reproduced verbitim from the NELS:88 student survey instrument in order to take advantage of the extensive field-testing of these items. The student-reported, background data that were collected includes: *education of the parents, household composition, educational aspirations,* and *prior grade retentions.* In addition, data were collected on a variety of parent and school-related conditions that might serve as potential intervening variables for explaining any statistical associations between program and outcome variables. Because the Americans All program included exercises to promote appreciative learning techniques in the classroom, involve parents in assignments, and stimulate student interest, additional data were collected on *parental involvement, students' perceptions of teachers and the school,* and *interest in selected subjects.* Most of these questions were also reproduced from the NELS:88 survey instrument.

Testing Procedures

Group testing procedures were used for both the Piers-Harris instrument (grades 3 and 5) and the Americans All instrument (grades 8 and 11). Classes were tested during their regular meetings on two separate occassions approximately four to five months apart. Students were assured that their answers would be kept confidential.

Statistical Analysis

With a pretest-posttest design with two nonequivalent groups, the analysis of program effects is most straightfowardly approached using a conditional or regression model of change (Plewis 1985). The simplest way of doing this is, first, to find the best-fitting linear relationship between the posttest and pretest scores of the variable for which one is seeking to observe change. Second, one computes the residuals, which is done by subtracting the actual posttest score from the score predicted on the basis of the linear relationship to the pretest. Next, one repeats the same two steps this time substituting an indicator variable that indexes membership in the program or control groups for the posttest variable. Finally, one regresses the posttest residuals against the program residuals. Since the resulting regression coefficent can be thought of as measuring the adjusted relationship between the posttest score and program indicator after removing their respective linear relationships with the pretest score, the size of the coefficient can be interpreted as an indicator of programmatic change. The larger the coefficient, the more change can be attributed to the program. Of course, this simple adjustment procedure assumes that no other important differences remain between the program and control groups which could affect the outcome. Otherwise, one would need to introduce these influences beforehand into the statistical adjustment of the posttest and program variables.

The above steps can be summed up in the following multiple regression equation:

$$Y_2 = \alpha + B_1 Y_1 + B_2 M + B_3 Z + e_2, \qquad (1)$$

where $Y_2$is the posttest score, Y_1 is the pretest score, $M$is a indicator variable that indexes membership in the program or control group, and Z is all other important influences of Y_2 which are correlated with M. If the error term e_2 satisfies the necessary statistical assumptions, the ordinary least squares estimate of B_2 will provide the best linear estimate of the program's effect.

Results

Sample Characteristics

Although regression techniques can be used to adjust statistically for prexisting differences between nonequivalent groups, it is also important to consider the consequences of attrition for the study's *external validity* (the representativeness of the final sample) and *internal validity* (the comparability of the program and control groups).

There are two sources of attrition that can potentially bias results: (1) attrition due to the non-participation of randomly assigned units, and (2) attrition due to the failure of participating units to complete the study. Typically, little data are available for assessing the first source of bias. In this study, for example, all that was known previously to the researcher about the assigned classrooms was the school, grade, and gender of the teacher. At assignment, the distributions for the classrooms were as follows: third grade (18%), fifth grade (18%), eighth grade (26%), eleventh grade (38%), and female teachers (62%). As the study progressed, however, third and eleventh grades tended to drop out more often than the middle grades. As a result, the grade distribution for the remaining classrooms in the study was more heavily weighted toward the fifth (40%) and eighth (40%) grades than the third (10%) and eleventh (10%) grades. The proportion of female teachers, however, remained approximately unchanged (60%).

The heavy attrition of third and eleventh graders obviously narrows the study's generalizability. But what about the representativeness of the fifth and eighth grade classrooms that remained in the study? A useful source of data for addressing this question is the NELS:88 national survey of eighth graders. One will recall that many of the questions in the Americans All upper-school questionnaire originally came from the NELS:88 survey. Table 1 compares the characteristics of a subsample from NELS:88 with the pretest and final sample of eighth graders

Table 1-- Characteristics of Eighth Graders

	NELS:88	Pretest Sample	Final Sample
Sample N	200	203	88
% Female	52.4	58.4	51.1
% African American	65.8	85.6	89.3
% Living with mother only	38.1	43.6	40.1
% College graduate (parents)	19.3	33.8	34.5
% Repeat a grade	28.9	29.3	33.0
Age		13.8	13.8
Self-esteem (Rosenberg)	2.6	2.6	2.7

in the study. The NELS:88 subsample was restricted to urban, public schools in the northeastern United States with minority enrollments of 90 percent or more. This is as close as one can come to matching the NELS:88 sample to the average profile of the Washington, D.C. public school system.

The data show that while the NELS:88 subsample includes a lower concentration of African-American students than the Americans All's sample of eighth graders, they are otherwise quite comparable in several important respects. Almost equivalent proportions of students lived with their mother only (around 40%). A lower proportion of the NELS:88 subsample had college-educated parents (19% vs 34%), but similar proportions of students had ever repeated a

grade (29% to 33%). The Rosenberg self-esteem scores were also nearly eqivalent, averaging between 2.6 and 2.7. Despite the severe attrition in the Washington, D.C. study, it appears that the sample of eighth graders who remained in the study does not look too dissimilar, with the exception of race, to eighth graders generally in northeastern U.S., urban public schools with 90 minority enrollments or more. Even if this were not the case, most studies can tolerate some loss of representativeness without biasing the comparision of program and control groups insofar as the attrition is similar in both groups. On the other hand, if the loss of cases differs between the two groups, serious problems of comparability can arise.

As already noted, the middle grades tended to remain in the study more frequently than the third and eleventh grades. This pattern characterized both program and control groups, but unluckily the loss of eleventh graders from the control group was complete. By the end of the data collection period, no juniors in the control group had completed the posttest. For all practical purposes, therefore, the upper school sample is a sample of eighth graders. Fortunately, the loss of third graders was less fateful. Roughly equal percentages remained in both the program and control groups.

The gender distributions at the upper-grade levels stayed approximately equivalent, but imbalances at the lower-grade levels could pose a problem. Whereas 64 percent of the third and fifth graders in the program group were female, only 40 percent were female in the control group. While prior studies show few sex differences in self-esteem scores using the Piers-Harris scale, girls and boys mature at different rates which could bias comparisons. This potential selection-maturation bias needs to be examined before drawing any conclusions about the program's effectiveness.

Data on racial and ethnic self-identification are available only for the upper school sample. The data show that classrooms with larger concentrations of African-American students were more likely to complete the study. The slight racial imbalance in the composition of the program and control groups at pretest became even more uneven at posttest. As a consequence, control group data are available only for African-American students.

A convenient way to summarize the compositional differences between the program and control groups is to compute a correlation matrix. The smaller the group differences are, the closer the correlation coefficients will be to zero. Large differences ($r > 0.150$) indicate areas of non-comparability. Table 2 shows that the lower school groups differ significantly by gender and the upper school groups differ significantly by grade and race. The signs of the correlation coefficients indicate that in the 3rd and 5th grades males are under-represented in program group. In the 8th and 11th grades, juniors are under-represented in the program group and blacks over-represented. For the other variables, the correlation coefficients are sufficiently close to zero so that the differences are ignorable.

In the following sections, I report the results of the regression analyses of the effects of Americans All on changes in student's self-esteem, cultural identity, and multicultural awareness. My first pass at the data assumes that the program and control groups are equivalent. No other variables aside from the pretest score and program indicator are included in the regression model. The program's coefficient in this regression model can be interpreted as an estimate of the program's effect under ideal experimental conditions. Since these ideal conditions were not achieved in practice, my second pass at the data introduces the appropriate control variables of gender, race and grade to adjust for pre-existing group differences. Because none of the control

Table 2-- Correlation Matrix

	Program	Male	Age	Grade	Repeat Grade	Black	College Parents
Program (Grades 3 & 5 ----------> 8 & 11)		*-.246*	-.096	-.038			
Male	.022		.069	-.036			
Age	.042	*.195*		*.855*			
Grade	*.175*	.107	*.681*				
Repeat grade	-.113	.170	*.517*	.127			
Black	*-.270*	*.183*	.143	.070	.161		
College parents	.117	-.121	-.103	-.007	*-.196*	*-.173*	
Mother only	.004	-.088	.099	.028	*.174*	.141	-.140

Italics: $p > .05$ (one-tailed test)

variables acted as suppressors (that is, obscuring program effects that were actually significant), I report the multivariate findings only for those program effects that were statistically significant in the first model.

Self-Esteem

Table 3 presents the results from the regression analysis of the effects of the Americans All program on children's self-esteem. Only the Piers-Harris scale was administered in the third and fifth grades. Both the Piers-Harris and the abbreviated Rosenberg scales were administered in

the upper grades.

The results are highly consistent across grades and instruments. There are virtually no statistically significant differences between the program and control groups. Once comparisons are adjusted for pretest scores, students assigned to the Americans All program evinced no greater improvement in self-esteem relative to students in the control group. In fact, the American All's standardized (beta) coefficients are slightly negative for the Pier-Harris scale. While the standardized coefficent is slightly positive for the Rosenberg scale, its size would have to be twice as large to be considered practically and statistically meaningful. The introduction of additional controls for gender, race and grade does little to alter this conclusion.

Table 3. Self-Esteem Coefficients

	Grades 3 and 5		Grades 8 and 11			
	Piers-Harris		Piers-Harris		Rosenberg	
Variables	Beta	p-value	Beta	p-value	Beta.	p-value
Americans All	-.106	0.184	-.017	.413	.122	0.13
Pretest Score	.504	0.005	.708	.000	.197	0.04
N	58		91		90	
R^2	28.9%		50.2%		6.9%	

Cultural Identity

Table 4 presents the results from the analysis of the program effects on children's sense of cultural identity. The questions were asked only of the upper grade students. Again the data show no significant improvements relative to the control group. In both cases the differences are in the expected directions: (1) a small drop in identification with popular cultural accomplishments (acting, sports, etc.) in favor of scientific, historical, religious and academic achievements, and (2) an increase in students' sense of their respective ethnic group's contributions to this country's history. Still, the magnitudes of the coefficients are much too small to warrant serious attention.

Table 4. Cultural Identity Coefficients

	Popular Accomplishment Scale		Own Group's Contributions to History	
Variable	Beta	p-value	Beta	p-value
Americans All	-.050	0.328	.146	0.126
Pretest Score	.379	0.001	.312	0.008
N	75		60	
R^2	14.9%		14.0%	

Multicultural Awareness

Table 5 presents the results from an analysis of the program effects on children's multicultural awareness. In this instance, the program effects are strong and statistically significant. Students in the Americans All program exhibited sizable differences in their appeciation and acceptance of cultural pluralism as compared to students in the control group. On average, there was an approximately 30 percent standard unit difference in Americans All participants' appreciation and acceptance of cultural pluralism. This corresponds to a 1.04 unit difference on the appreciation scale (posttest $\bar{x}$ =7.42, s.d. = 1.69) and 0.85 unit difference on the acceptance scale (posttest $\bar{x}$ = 4.54, s.d = 1.41).

Figure 2 gives a visual display of the findings for pluralistic appreciation. It is a scatterplot of the posttest scores against the pretest scores. The two lines show the pooled within-group regressions. The lines differ by a constant of 1.04 units which equals the unstandardized program coefficent. It can be interpreted as the estimated effectiveness of Americans All in promoting students' appreciation of pluralistic values. For fixed values of the pretest, program participants scored, on average, one unit higher on the posttest than students in the control group. This corresponds to a 30 percent standard unit gain.

The results in Figure 2 were obtained by constraining the regression analysis to fit parallel lines. Closer inspection of the data points suggests, however, that the within-group regressions might not be parallel. This possibility can easily be accomodated in the regression analysis by allowing for a group-pretest interaction term. Given the small size of the sample, the interaction term is unlikely to achieve statistical significance. But if it is included, then the group regressions cross at the lower end of the scale. Figure 3 illustrates the effect: the average gain declines as the pretest score becomes lower and rises as the score becomes larger.

Table 5. Multicultural Awareness Coefficients

Variable	Pluralistic Appreciation		Pluralistic Acceptance		Other Group's Contributions to History	
	Beta	p-value	Beta	p-value	Beta	p-value
Americans All	.302	.005	.281	.015	-.101	.197
Pretest Score	.424	.000	.140	.134	.547	.000
N	59		60		54	
R^2	26.2%		10.3%		27.4%	

There are several ways to handle nonparallel regressions. If one assumes that the two groups form a simple random sample from a population, then Rubin (1977) provides a formula for estimating the average group effect over all values of the pretest in the population. This works out to be 1.00 for the sample of Washington D.C. eighth graders. Because it is roughly equivalent to the previous unstandardized estimate, it seems that little is gained by relaxing the assumption of parallel regressions.

By comparison, there was no relative increase in students' sense of the contributions of other ethnic groups to American history. In fact, there was a small but statistically insignificant decrease. This downward shift is related, in part, to the slightly lowered ratings of the historical contributions of European Americans by Americans All participants (see Table 6). The data suggest

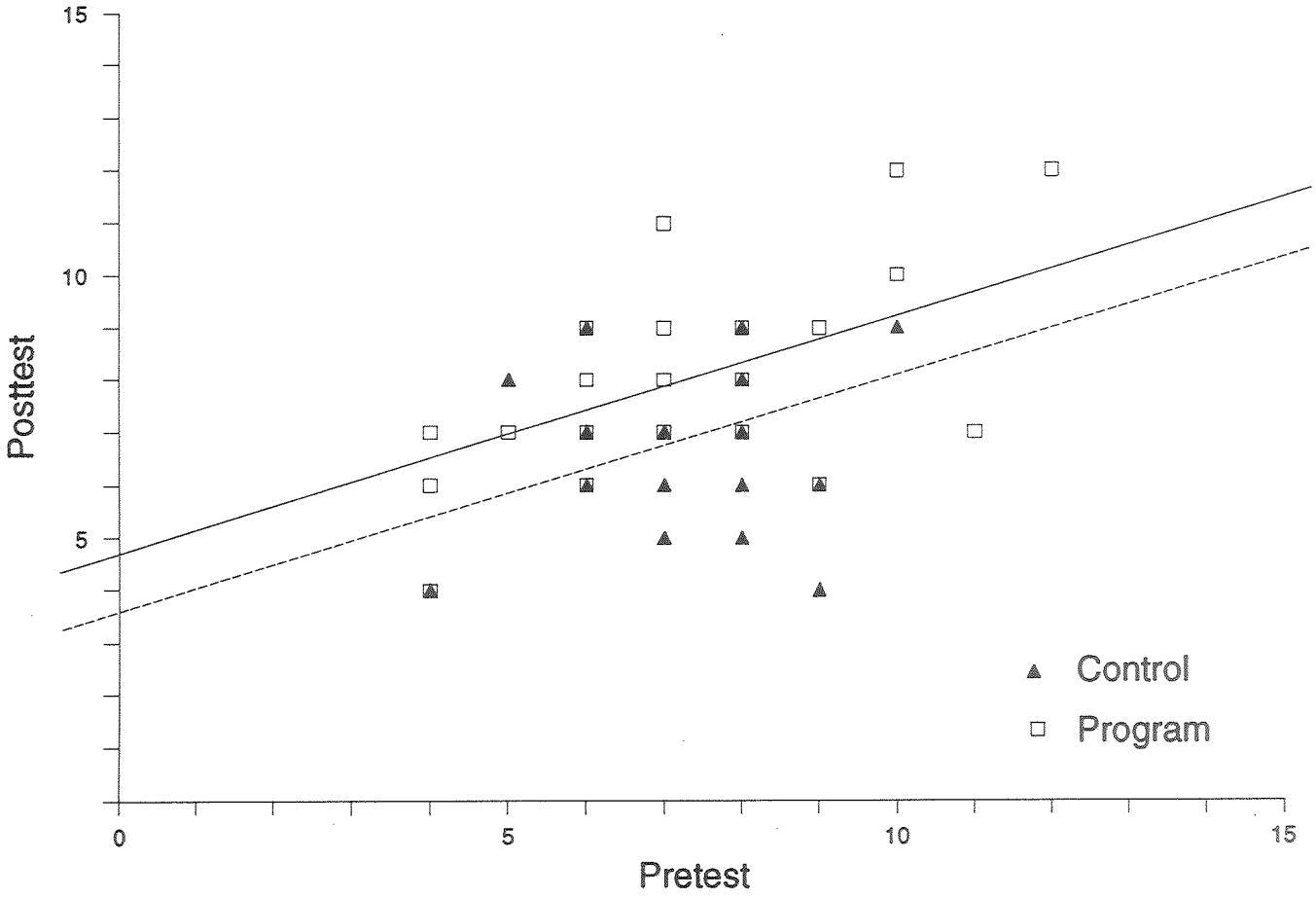

Figure 2. Plot of pluralistic appreciation posttest against pretest. The lines give the pooled-group regressions. Program: ———Control: ----------

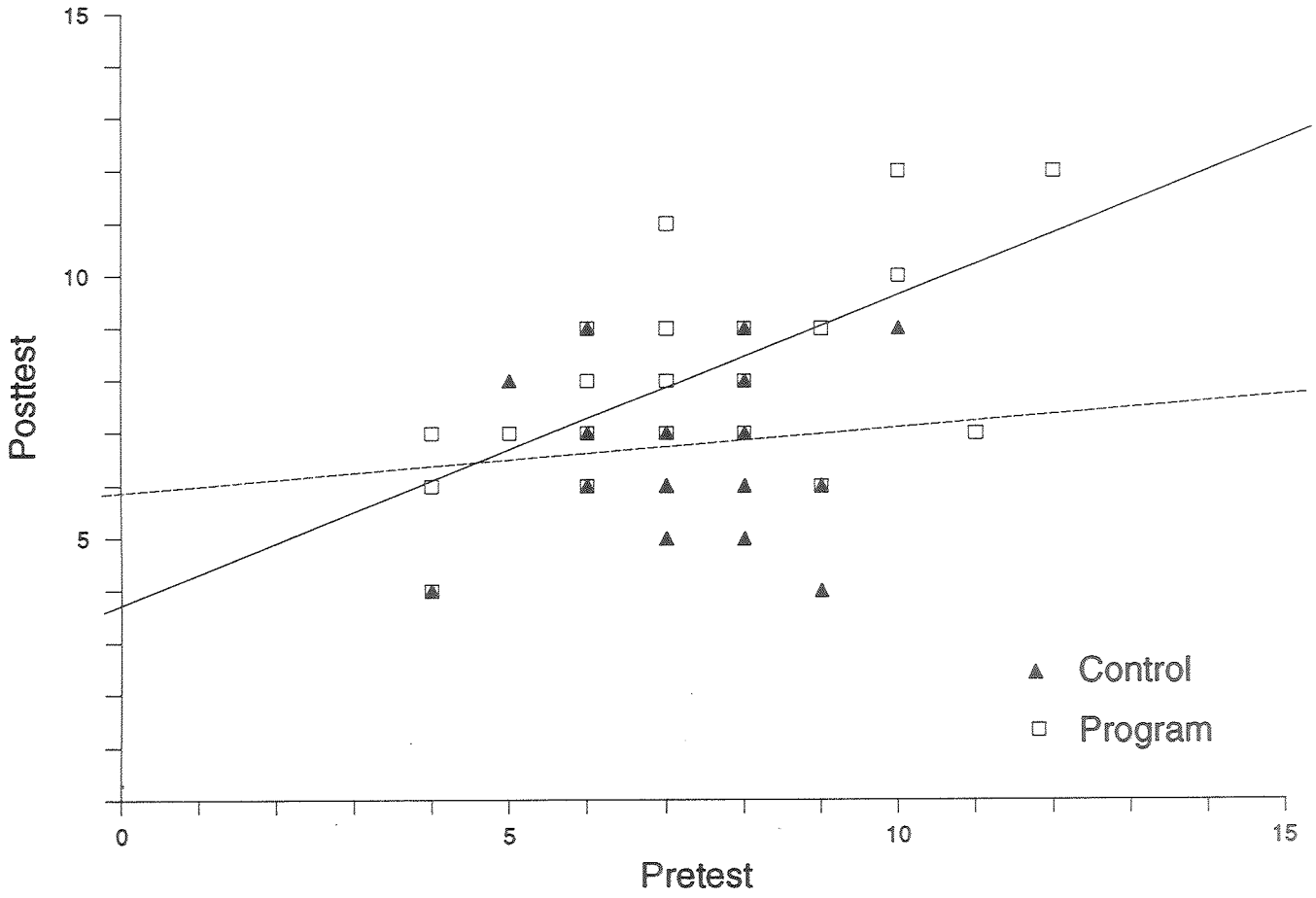

Figure 3. Plot of pluralistic appreciation posttest against pretest. The lines give the within-group regressions. Program: ———Control: ———

Table 6. Ethnic Contributions to American History

Variable	African American Beta	African American p-value	European American Beta	European American p-value	Mexican American Beta	Mexican American p-value
Americans All	.227	.032	-.197	.062	-.088	0.465
Pretest Score	.346	.003	.266	.020	.341	0.005
N	61		60		60	
R^2	20.8%		9.9%		11.5%	

	Puerto Rican American Beta	Puerto Rican American p-value	Asian American Beta	Asian American p-value	Native American Beta	Native American p-value
Americans All	-.021	.433	-.029	.406	.067	.287
Pretest Score	.389	.001	.478	.000	.478	.000
N	58		57		60	
R^2	15.1%		23.4		24.6	

that the Americans All program helped to boost the ratings of African Americans' contributions, but lower slightly the ratings of European Americans' contributions. There was no apparent relative change in the ratings of the other ethnic groups.

Additional Statistical Adjustments

It will be recalled that attrition resulted in the complete loss of juniors and non-blacks from the control group. Would the results on pluralistic appreciation and acceptance have turned out differently if the program and control group had included equivalent proportions of juniors and white and Hispanic students? The data on the program participants show that juniors and non-black students exhibited larger gains in multicultural awareness than did 8th graders and black students. This fact alone does not invalidate the comparison unless it could be demonstrated that juniors and non-blacks would have achieved these same gains on their own without the Americans All program. Although there is little reason to believe that mulitcultural awareness develops differently by age and race, the absence of juniors and non-blacks in the control group does weaken one's confidence in the validity of the findings.

A simple check on the robustness of these findings is to include race and grade indicators in the regression model. This helps to separate out the changes that are attributable to being a junior or white and Hispanic from the changes that are attributable to being in the program. Making this statistical adjustment reduces the Americans All (beta) coefficients from .30 to .23 for pluralistic appreciation and from .28 to .22 for pluralistic acceptance. Both adjusted coefficients remain statistically significant at the .10 level. Another check is to restrict the comparison only to black students in the 8th grade. The results are similar. Still, one needs to be cautious in drawing inferences. Even though the program effects do not entirely wash out with the introduction of controls or with narrower sample restrictions, the small sizes of the program coefficient

does point to the need for replicating the study with a larger and more diverse sample of students.

Discussion

Americans All is a national multicultural education program that strives to promote appreciation for self and sensitivity to cultural diversity by drawing on the history of voluntary immigration to the United States and the involuntary incorporation of various ethnic and racial groups through colonization, enslavement, or territorial annexation. The results of this study show that Americans All can be effective in promoting students' sensitivity to cultural diversity. Students who participated in the program exhibited significant gains relative to the control group in their approval of pluralistic values and in their acceptance of foreign-born persons.

The results on enhancing children's self-esteem and reinforcing their sense of cultural identity are less promising. There is little evidence of a significant program impact. One possible reason for the lack of change in students' self-esteem is that the children's scores were already high at the outset. Hence, there was little room for promoting gains. Another is consistent with the idea that self-concept is a relatively stable trait (Blascovich and Tomaka 1991). Like other person traits (e.g. intelligence, Type A behavior), self-esteem cannot readily be manipulated experimentally. Even when dealing with young children, it may be difficult to induce measureable changes in a limited time frame when children are assessing themselves against 8 to 13 years of self-evaluative experiences. This inability to manipulate self-concept experimentally obviously poses difficulties for researchers interested in evaluating interventions designed to raise self-esteem. One possibility is to extend the period of intervention in the hopes that change becomes more observable with time. Another, suggested by Blascovich and Tomaka (1991), is to focus on self-evaluations of very specific or novel attributes. For example, specific

questions, such as "I wish my eyes were blue" or "I am knowledgeable about foreign lands", might be better for assessing the impact of multicultural programs than global ones, such as "I am a happy person" or "I do many bad things."

Cultural identity can present similar difficulties. Asking students to assess the historical contributions of their own ethnic group shows few differences between the program and control groups. It is possible that the question is too general or vague to register the effect of the program. Questions that are geared to specific content, such as "Where is the oldest American university located?" (Answer: Mexico City) or "Who performed the first sucessful open heart surgery?" (Answer: Daniel Hale Williams, a black surgeon) might be better for gauging children's awareness of the historical contributions of members of their own ethnic groups.

Some further refinement of instruments is clearly in order. Still, in light of the weak showing of these variables, it is worth asking how essential to the mission of multicultural education is improving children's self-concepts and strengthening their cultural identities. If one defines that mission to be the promotion of cultural pluralism and tolerance for other people's differences, the results of this study suggest that significant gains can be made without achieving corresponding changes in self-esteem or cultural identity. But if one takes the mission to be improving the school performance of minority ethnic and racial groups, then clearly more in the way of bolstering stuident self-confidence is required.

Multicultural Education and Academic Achievement

Several explanations of ethnic variation in school performance point to the possibility of encouraging improvements through multicultural education. One is the hypothesis that racial and ethnic stereotypes undermine minority children's self-confidence for school success (Ausubel and Ausubel 1963, Kvaraceus 1965, Witty 1967). Multicultural education can help to

counter this erosion of self-concept by neutralizing negative stereotypes. But as mentioned previously, this intervention strategy is called into question by research showing that minority children have no lower or perhaps higher self-esteem than majority white students (Crocker and Major 1989, Hoelter 1983, Rosenberg 1979).

Another explanation is the theory of self-fulfilling prophesies: minority children do poorly in school because teachers do not expect them to succeed. According to this view, minority students are assigned to remedial classes and inferior schools by means of misclassification and biased testing. Eventually, they may also come to behave in ways that are consistent with these evaluations (Miller and Turnbull 1986). Multicultural education can help to break this self-fulfilling cycle by sensitizing teachers to the cultural biases and prejudices they may bring to the classroom. Although this explanation is generally accepted, Ogbu (1991) criticizes it as failing to explain why immigrant students who attend the same inferior schools do relatively better than native-born blacks and other minorities. For the same reasons, he criticizes the theory of cultural discontinuities which posits that school failures are related to cultural conflicts between teachers and students in styles of instruction and learning. This theory, he notes, also falls to account for why immigrant students with significant language and cultural differences often perform better than native-born students of the same ethnic origins.

Ogbu (1990) argues for a distinction to be draw between minority students whose presence in the United States is the historical result of voluntary immigration and minority students whose presence is the historical result of involuntary incorporation through colonization, territorial annexation or enslavement. Most descendents of European immigrants and many from Asia fall into the former grouping, while most Native Americans, African Americans, Puerto Ricans, and southwestern Mexican Americans fall into the latter. Ogbu says that immigrant students tend to differ from involuntary minority students in their orientations toward the cultural and language

discontinuities they encounter in school. Immigrant students define these discontinuities as barriers to be overcome to achieve desired education, while native-born minorities tend to define these differences as markers of their cultural identity that must be maintained. Ogbu says that under these circumstances:

> "it is not enough for teachers and schools to know a good deal about student's cultures and languages and to use the knowledge in designing school curricula or in teaching. A true cultural diversity that promotes the academic success of minority students and other marginal populations is one that permits them to cross cultural and language boundaries without feeling threatened." (Ogbu 1990, p. 428-29).

For example, as Ogbu (1991) notes, many young African Americans view sports and entertainment, rather than education, as the way to get ahead. Their perceptions are reinforced by first-hand observation and by a media that glorify black atheletes and entertainers but publicize the academic failings of blacks and other minorities. As Crocker and Major (1989) observe, members of marginal populations gradually come to regard those dimensions on which members of their group excel as more central to their self-concept than those dimensions on which their group fares poorly. Academic achievement is related to self-esteem only to the extent to which academic experiences are central to one's self-concept. It is little wonder then that low-acheiving, inner-city students can score higher in global self-esteem than their higher-achieving middle-class counterparts (Jordan 1981). The implication for multicultural education is that fostering student self-esteem may not be sufficient for promoting school success. What is needed is the creation of a supportive school environment that enables minority children to invest more of their self-concept in academic success without feeling threatened or vulnerable to ostracism by their peers. In their ethnographic study of Washington, D.C. high school students, Forham and Ogbu (1986) reported that black students regarded many behaviors associated with

high academic achievement as "acting white." Black students who wanted to achieve felt pressured to downplay their academic interests and to engage in peer-approved activities that jeopardized their scholastic standing. As Ogbu notes, they face a conflict that does not seem to confront immigrant students between "striving for academic success and maintaining their minority identity and cultural frame of reference" (Ogbu 1991, p. 536).

The Americans All questionnaire for the upper grades included a question: "What are the things about your ethnic or cultural group that make you feel most proud?" Black students less often selected scientific, historical, religious, and academic achievements ($\overline{X} = 5.8$) than white and Hispanic students ($\overline{X} = 6.6$). One way that multicultural education can assist students in crossing cultural boundaries without fearing the loss of cultural identity is to acquaint them with the scholarly and scientific contributions of their ethnic forbears. For example, the Americans All program highlights the scientific, literary, and historical contributions of members of various ethnic groups to American culture. Although this study revealed no sigificant change in students' attitudes toward such accomplishments, futher investment in exploring student sentiments and feelings about cultural identity might help to provide more secure transitional bridges for crossing perceived cultural boundaries.

Challenges of Multicultural Education

As social studies become more diversified, the history of racial and ethnic subordination in this country will become a more prominent feature of the school curriculum. Some people will welcome the change as finally balancing out the scales of historical interpretation. Others will no doubt worry about tipping the scales too far in the direction of an overly ethnic intepretation of American history. Whether one views the coming change as equilibrating or not, it will nonetheless serve as a painful reminder of the terrible compromises the nation has made in extending

the rights of life, liberty, and equality to its people. Because many of these past injustices are currently visible in the plights of various racial and ethnic groups today, issues of collective fault and reponsibility will inevitably be raised. Feelings of anger and guilt will be aroused. Multicultural educators must be prepared to deal with these issues and feelings constructively.

Conclusion

This study demonstrates that a program of multicultural education can be helpful in promoting student appeciation of and tolerance for cultural differences. These are important accomplishments in light of recent incidents of ethnic intolerance and racial violence in our nations cities and schools. As the country grows more culturally diverse, encouraging appreciation and acceptance of this diversity will become an increasingly important component of citizenship education in the United States.

The study also raises some questions about the need for and capacity of multicultural programs to raise children's self-esteem. Previous research has shown the self-esteem of minority students to be no worse and perhaps slightly better than the self-esteem of non-Hispanic white students. Perhaps more sensitive instruments for assessing the self-esteem of ethnically diverse populations need to be devised. In any event, by helping to create a non-threatening classroom atmosphere that enables minority children to invest more of their self-concept in academic success without seeming to reject their minority identity and cultural frame of reference, multicultural education offers another promising avenue for promoting educational success.