3 Brook

DESIGN FOR A NATIONAL LONGITUDINAL STUDY OF SCHOOL DESEGREGATION: VOLUME II. RESEARCH DESIGN AND PROCEDURES

PREPARED FOR THE U.S. COMMISSION ON CIVIL RIGHTS

R.L. CRAIN, D. ARMOR, F.G. CHRISTEN,
N.J. KING, M.W. McLAUGHLIN,
G.C. SUMNER, M.A. THOMAS, J.J. VANECKO

R-1516/2-USCCR
SEPTEMBER 1974

Rand
SANTA MONICA, CA. 90406

The work upon which this publication is based was performed pursuant to Contract CR3AK009 with the U.S. Commission on Civil Rights. Views or conclusions contained in this study should not be interpreted as representing the official opinion or policy of the U.S. Commission on Civil Rights.

Published by The Rand Corporation

Photo courtesy THE NEWS AMERICAN, Baltimore, Maryland

DESIGN FOR A NATIONAL LONGITUDINAL STUDY OF SCHOOL DESEGREGATION: VOLUME II. RESEARCH DESIGN AND PROCEDURES

PREPARED FOR THE U.S. COMMISSION ON CIVIL RIGHTS

R.L. CRAIN, D. ARMOR, F.G. CHRISTEN,
N.J. KING, M.W. McLAUGHLIN,
G.C. SUMNER, M.A. THOMAS, J.J. VANECKO

R-1516/2-USCCR
SEPTEMBER 1974

Rand
SANTA MONICA, CA. 90406

PREFACE

The U.S. Commission on Civil Rights asked The Rand Corporation to prepare a design for an exhaustive research program on school desegregation—research to answer a number of questions about school desegregation and to locate ways in which policy-makers can intervene to make desegregation work more effectively. This is Vol. II of the final report. Volume I contains a summary of the report, and a lengthy discussion of the issues involved in designing this research program. (The summary has also been published separately for readers interested in that alone.) This volume is a detailed discussion of the research design itself. In addition, appendices containing questionnaires and supplemental notes on sampling are available.

Volume II is intended for those government officials and professional researchers who are involved in developing research on school desegregation. It is not intended to stand alone; reading the summary and portions of Vol. I is necessary for an understanding of the issues being dealt with here.

The 11 sections of this volume describe a general design for a research program in school desegregation. The research program consists of nine different studies, many of which share a common data base. In Vol. I we describe the issues that have led us to the research focus and methodology used in this design.

Briefly, two main factors led us to the design developed here:

- The Commission on Civil Rights asked for a research design covering most of the major policy issues in school desegregation. After reviewing the literature and meeting with a number of policy-makers and researchers, we concluded that the range of policy concerns is very great, and that there was very little agreement as to which policy issues were most important. It therefore became necessary to develop research to deal with all of these policy issues.
- We concluded that the intense controversy surrounding existing research on desegregation meant that no single study could succeed in settling the outstanding issues. We therefore designed a set of studies which used different methodologies and could be conducted by different researchers so that the interchange between the studies might lead to a consensus about the major conclusions.

The result is a program of studies designed to answer a number of different questions about school desegregation as it applies to four groups: Puerto Ricans, Mexican-Americans, blacks, and Anglo-Americans. (The Commission had earlier decided that it would be impossible to also deal with Cuban immigrants, American Indians, Americans of Oriental ancestry, or other ethnic groups within a single research program.)

The first five sections of this volume (6 through 10) deal with the six main studies, which share a common sample. The sample, described in detail in Sec. 10, can be thought of as a funneling process.

First, a screener survey of 750 school districts is used to locate 220 districts representing a wide range of desegregation experiences. This sample is used for the community reaction survey described in Sec. 6, a study of the politics of desegregation done in order to determine how districts might go about desegregating with the least cost in terms of loss of support for the schools and with the greatest acceptance of desegregation by minority and majority citizens.

Second, a sample of 1600 elementary schools, junior high schools, and high schools is selected. In Sec. 7 we outline a one-year longitudinal study of student attitudes and achievement in these schools. This study is designed to test a very wide range of hypotheses about the effects of school desegregation on students and the impact of school policies on success of desegregation.

In Secs. 8 and 9 we describe four substudies, which follow the large-scale study; each takes a subsample of schools and pursues a particular topic more intensively. The student panel study follows students for two additional years in order to give a longer-term portrait of school effects. A second study follows newly desegregated schools to observe how they adjust to desegregation and to discover policies which will assist them in doing so. Finally, the studies of innovative elementary schools and innovative secondary schools take particular innovations and study them in detail in an effort to locate policy instruments of value. The large-scale study provides preliminary hypotheses and data from which to draw the special samples for these four studies.

After Sec. 10, which discusses the sample, are four sections dealing with other studies. Section 11 is a design for a study of the impact of desegregation policies on the exodus of whites from central cities. Section 12 discusses how the studies of innovative schools can be converted to the more powerful methodology of the randomized experiment. Section 13 discusses how three existing panels of young adults can be used to test hypotheses about the impact of different types of desegregation on college attendance, job-hunting behavior, and adult racial behavior. Section 14 proposes a series of social-psychological laboratory experiments to test hypotheses about the dynamics of race relations in schools.

Finally Secs. 15 and 16 discuss general problems of the program. Section 15 outlines the cost of the study and suggests modifications which might be made to reduce costs, and Sec. 16 is a general discussion of the overall management of the program.

CONTENTS

B. Brooks

6. COMMUNITY REACTION TO DESEGREGATION

In this section we propose the first study: a survey of how communities have dealt with desegregation. The study design is based on the model outlined in Sec. 4 of this report (Fig. 4.1). Briefly the community reaction study has six concerns:

- Techniques school districts have used to desegregate schools.
- Reactions of community leaders, organizations, and the media to those techniques.
- How different desegregation techniques and different reactions by community leaders influence the reactions of black, Puerto Rican, Mexican-American, and Anglo parents.
- Parent attitudes toward desegregation in terms of their social background and general social ideology.
- Parent fears about their children's safety in desegregated schools.
- Social and political effects of desegregation on the community.

This six-step analysis is intended to tell us ultimately what steps federal, state, and local leaders should take to maximize citizen acceptance of desegregation and maintain the economic and political support of the community.

The study requires three different types of data collections, repeated at various points to create a longitudinal data set.

A *screener survey,* to determine the extent of desegregation in a large number of districts; a *district leadership survey,* to determine the details of the desegregation plan and its implementation (or the details of how the community failed to adopt a desegregation plan) and also to determine the reaction of community leaders; and a *parent survey,* measuring attitudes toward the desegregation plan and degree of parent support for the schools. Data are collected over a five-year period to permit a longitudinal analysis of community reaction.

This section first desribes the samples and then presents the data collection methods and the analysis plan.

SAMPLING

Each of these three data collection efforts requires a different sampling strategy. The same samples are used for the other studies in the research program in order to reduce costs and permit the studies to exchange data. The sampling scheme is complex and is described in detail in Sec. 10 and Appendix B. Briefly, the three samples used are as follows:

The screener sample consists of 750 school districts, including 450 northern districts and 300 southern districts, all having either over 3000 blacks or over 3000 persons of Spanish heritage. The *district* sample consists of 220 school districts selected from the 750. The school sample is constructed by using these 220 school districts as a universe from which to sample (approximately) 548 elementary school attendance areas. A sample of 15,000 parents of third-grade students in these schools is used in the community reaction study.

Until the sample is drawn, the exact number of schools, students, or parents is not known; approximate figures are used in this report.

For the purpose of this section, it is necessary to point out the following characteristics of the sample:

The screener sample includes all northern metropolitan school districts with a significant number of minority residents, and approximately half of all southern districts with a significant number of blacks. Thus this sample is large enough to encompass virtually every technique of desegregation.

The district leadership sample is drawn with pairs of districts matched by community social characteristics. In half the cases the pairs will also be matched by degree of desegregation, permitting us to consider the differences between similar districts that desegregated, but perhaps in different ways. In the other half the two districts will differ in degree of desegregation, permitting us to ask why two otherwise similar districts should differ in their decision to desegregate.

The district sample is drawn so as to oversample northern court-ordered desegregation, which is relatively rare. Highly desegregated districts are also oversampled. The sample includes all regions and Mexican-American and Puerto Rican as well as black and Anglo students.

The sampling of parents is based on a sample of schools, and until that sample is drawn, we cannot know how many districts are included. For the sake of this presentation, we will assume 160 districts; the true number will be between 125 and 200.

The elementary school attendance areas are matched, so that we can compare students of one ethnic group in a desegregated school with students of similar social backgrounds in segregated schools. This means that we can compare parents whose children are in desegregated schools with parents of the same ethnicity and similar social background whose children are in segregated schools.[1]

The decision to sample parents means that the study will not obtain data on attitudes of nonparents, who could of course be added, but we do not believe the cost is justified.

THE LONGITUDINAL ANALYSIS

Figure 6.1 portrays the way in which the three types of surveys are combined and repeated to produce the data for a five-year longitudinal analysis. There are eight data collection steps, beginning with a broad survey of a large number of districts, gradually funneling down to detailed analysis of a small number of districts in the process of desegregation. The timing of several of the steps will permit the data to be used in other studies as well as in the community reaction survey. The eight data collection steps, with the analysis intended for each, are outlined below.

Step 1: The Initial Screener Survey

The screener survey is a solution to the conflict between the need for a large sample and the desire to keep costs within practical limits. The data in this part of the study can be easily collected through mail questionnaires and phone interviews, from the U.S. census and from HEW directories. School desegregation action, school integration levels, and demographic characteristics are the three types of data that will be collected in the 750 school districts.

[1] A number of other constraints on the sampling of both school districts and schools are presented in Sec. 10.

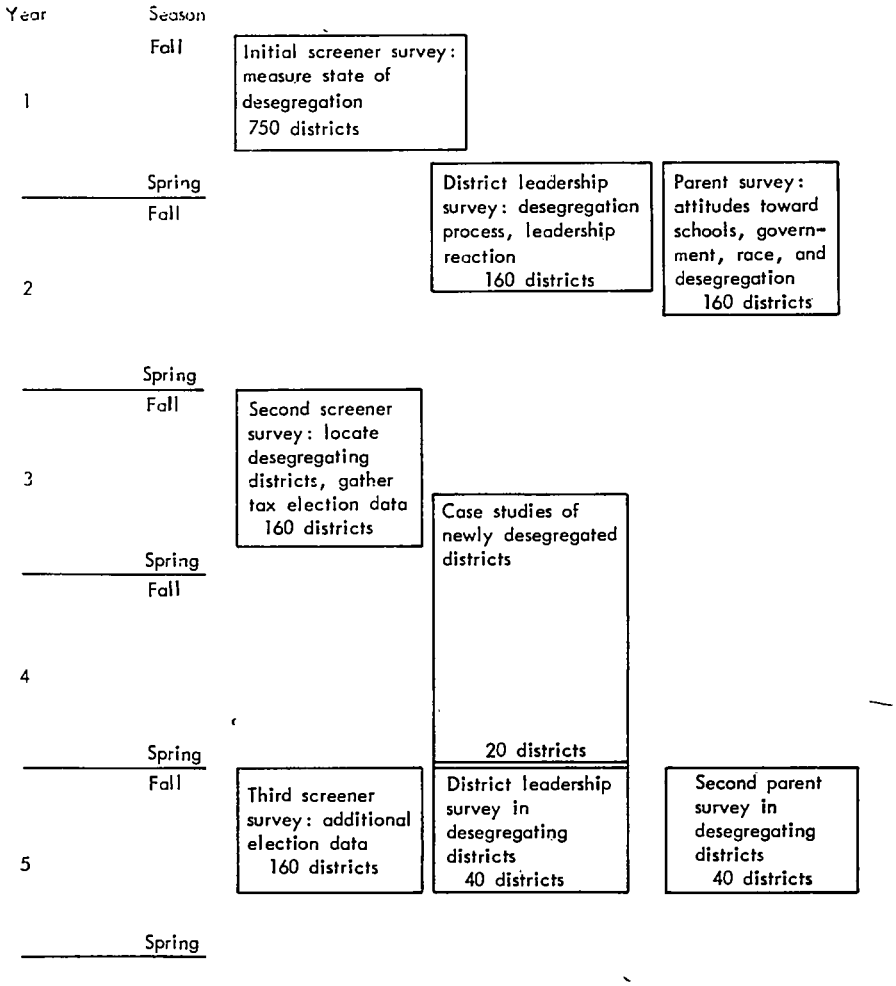


Fig. 6.1—Data collection plan for the community reaction study

The most important data to be collected are those on school desegregation action. The technique used will be one developed by Rossell and Crain (Kirby et al. 1973), explained in detail in their Chapter 12. A mail questionnaire will be sent to 750 school districts listing their biracial schools. Administrators will be asked to indicate on the questionnaire why each school is biracial by checking the appropriate column. The school desegregation variable is then computed from HEW school racial census data by subtracting the racial composition of a school in the year of a desegregation action from the racial composition in the previous year. The difference would be noted (if it increased desegregation in the school) as a number of black and white students reassigned in that year for the purposes of integration. The characteristics of this desegregation action (e.g., voluntary versus mandatory) would be additional data. Initially this would involve a retrospective history. Hopefully a census of action to desegregate would be kept in the future as well. The source of racial composition data for the listing of schools is the HEW school survey for 1970. We anticipate the use of 1967, 1968, 1970, and 1972 directories and 1974 data tapes by the time this study is launched.

Most of the data can be obtained by mail questionnaires with telephone follow up. (When this was done previously, the return rate by mail was only 50 percent; telephone interviewing brought the rate up to 98 percent.) The questionnaires would yield the following data:

1. A measure of school desegregation action: number and percentage of black and white students reassigned to predominantly white and black schools, respectively, in each school and each year from 1967 to the year of study.
2. Number and percentage of schools in which desegregation actions have been taken in each year.
3. Characteristics of the desegregation: e.g., mandatory busing, voluntary busing, boundary changes, court order, high school, junior high, or elementary.

In addition, several measures of school segregation would be computed by using the HEW directory, including an index of dissimilarity (Taeuber and Taeuber 1965; Farley and Taeuber 1973). The dissimilarity index could be compared to other indices such as Crain's (in Kirby et al. 1973, Appendix G). These measures will show how much segregation is in a school or school district, but they will not distinguish between administratively initiated desegregation and that which is the result of residential integration in the attendance zone of a school. On the other hand, the measure of desegregation action does not show how desegregated a school or district is as a result of reassigning some number of its black students to predominantly white schools. Therefore, both types of measures are needed to draw the complete picture. Data would be obtained concerning characteristics of school desegregation litigation and state-imposed sanctions by means of a mail questionnaire to the legal officer of the school system.

Thus, for each of 750 school districts, the data in Chart I would be assembled into a file.

These data would be used mainly to permit sampling of the 220 districts for the district leadership survey. They will tell us which districts have desegregated, and also which districts are likely to desegregate in the next few years, either because of voluntary plans or anticipated court orders. In addition, these data will permit a thorough description of the state of school desegregation, and a census of the techniques used to accomplish desegregation.

Chart 1

DATA COLLECTED IN SCREENER SURVEY

Data	Source
Social, economic, and political characteristics of community	U.S. census, municipal yearbook
Taeuber index of school segregation and other indices	HEW Directory of Public Schools
Rossell index of desegregation action	Questionnaire to superintendent
Characteristics of litigation and state education department action	Questionnaire to legal officer

Step 2: The District Leadership Survey

The district leadership survey is our main tool for determining the details of the desegregation decision and its implementation, the details of any decisions not to desegregate, and the way in which the community leadership supported or failed to support the schools. Usually, these sorts of data are gathered entirely by case study methods; but case studies are by nature unsystematic and difficult to generalize. We propose to use case studies, but to support them with a set of comparative studies using systematic interviewing of community leaders.[2] The sample for this step is the district sample (220 districts). Costs have been reduced by using only approximately 160 districts in which schools are sampled (and hence parents interviewed). The data will come from mail questionnaires and telephone interviews of *positional elites,* i.e., persons occupying positions in which they either have influence on the school desegregation process or have access to information concerning desegregation. The principle behind the design of such a sample is to include the most important actors and the most informed sources while at the same time representing opposite biases in order to triangulate measures of important variables.

The seventeen or more different interviews contain a common core of questions about community controversy, influence patterns, general leadership reaction, and community electoral behavior. Some of them would have supplements (e.g., the civil rights leaders as well as the black, Chicano, and Puerto Rican political leaders would be asked some questions about their major concerns and demands on the school system and questions about internal organization and leadership of the minority communities). There will also be a supplement to all the questionnaires which will only be utilized in desegregated cities. Chart 2 illustrates the subsets of respondents and questionnaire forms utilized in different cities. Chart 3 outlines the data to be obtained from the basic form questionnaires.

The fifteen types of respondents represent an effort to include major types of positions—school officials, political leaders, persons from the media, community, and civil rights organizations—and to balance opposing viewpoints. The selection of two retired school board members is based on the likelihood that they will be both cooperative and highly informed and will have a more balanced view than someone now in public office. A final data collection effort—or more appropriately, coding effort—would involve using the Library of Congress microfilm newspaper collection, the *New York Times* computer accessible index and morgue, and the Bell and Howell Urban Research Corporation Update file to code level and type of controversy as suggested below:

1. Number of picture and column inches and time period devoted to the (a) school desegregation issue, (b) school desegregation demand, (c) school desegregation implementation.
2. Number of front page articles, number of editorials, and time period on issue, demand, implementation.
3. Size of crowds at school board meetings as reported in the newspapers (and time period).
4. Number of times school desegregation mentioned as campaign issue in mayoralty and councilmanic elections, school board and school financial referenda, and number of candidates taking a stand as reported in newspaper.

[2] See Rossi and Crain (1968) for a description of this method. A special issue of *Social Science Quarterly* on community decisionmaking and conflict (forthcoming) offers several examples of the method in practice.

Chart 2

LEADERSHIP INTERVIEWS INDICATING QUESTIONNAIRE SECTIONS TO BE ADMINISTERED TO EACH AND NUMBER OF SCHOOL DISTRICTS IN WHICH THEY WILL BE ADMINISTERED

Respondent	Number of Districts in Cell				
	Basic Form	Internal Organization of Minority Leadership and Priorities of Demands	Stages of Controversy and Implementation Procedures	Personal Views on Desegregation	Organization or Institution Response[a]
1. School Board President	160	0	90[b]	160	0
2. 1963 School Board President	160	0	0	160	0
3. Current Minority School Board Member[c]	200	200	100	200	0
4. Past Minority School Board Member[c]	200	200	0	200	0
5. Superintendent of Schools	160	0	90	0	0
6. News Director of Local TV Station (or Radio)	160	0	90	0	160
7. City Editor of the Largest Circulation Local Newspaper	160	0	90	160	160
8. PTA President	160	0	90	160	160
9. Mayor	160	0	90	160	0
10. Leader of Opposition Party (or Group) to Mayor	160	0	90	160	0
11. NAACP Director or Equivalent[c]	200	200	100	200	200
12. Highest Minority Elected Official[c] (excluding Federal)	200	200	100	200	200
13. Urban League Director or Equivalent[c]	200	200	100	200	200
14. Editor--Minority Newspaper[c]	200	200	100	200	200
15. Three Ad Hoc Group Leaders Living in Attendance Zones of Desegregated Schools	480	0	0	480	0

[a]This refers to the organizational or institutional position for which the respondent is selected. For example, there will be questions for the editor regarding the editorial positions of the newspaper.

[b]These pertain to the desegregation process and thus are only conducted in desegregated districts.

[c]In triethnic communities, both Black and Mexican-American or Puerto Rican leaders will be interviewed; we estimate that this will produce 200 interviews in the 160 districts.

5. Number and length of duration of boycotts, riots, sit-ins that have school conditions as focus as reported in newspapers.

A lengthy analysis of these data, using primarily multiple regression and path analysis methods, should provide partial answers to the following questions: How do various methods of presenting the school desegregation issue to the community affect the support or opposition of local leaders? What factors cause a school system to adopt a desegregation plan at all, and what factors influence the kind of plan they choose? What community factors cause a desegregation plan to be whole-heartedly implemented by school staff?

Step 3: The Parent Survey

The sample of 15,000 parents is drawn from the 548 elementary schools. This means that analysis can be conducted at the *school* level as well as the school *district* level. Three different types of data will be used in the school-level analysis: attitudinal and voting data from a survey of parents; the elite interview data described in the section above; data obtained from the screener survey also described above; and

Chart 3

OUTLINE OF QUESTIONNAIRES FOR LEADERSHIP SURVEY

Basic form administered to all respondents:

1. What has been public behavior of key officials on desegregation?
2. (From school officials) what are characteristics of desegregation plan?
3. What is character of public presentation of plan?
4. How much controversy over school desegregation?
   a. Prior to any action
   b. During court case (where applicable)
   c. After plan but before implementation (where applicable)
   d. During implementation (where applicable)
   e. After implementation (where applicable)
5. Number of public demonstrations for or against desegregation.
6. What public position have major organizations, institutions, and individuals taken (support-opposition)?
7. How serious a problem is school desegregation?
8. How serious a problem is school financial support and how is this related to school desegregation?
9. How is authority, power, prestige distributed in the system?
10. Estimate of who could have prevented problems of controversies, who could have provided solution.
11. How good a job is school board doing?
12. How good a job is school superintendent doing?
13. Is metropolitan school system a feasible solution?
14. Is busing a feasible solution?
15. What is general quality of schools?

voting returns. These will be supplemented with case studies of selected cities as described in Sec. 9.

In all but 80 of the 548 schools (see below), the sample will consist of 12 parents of third-grade children from each ethnic group represented in the school. This means that in a one-group school (for example, a segregated Anglo school) 12 parents will be interviewed; in a triethnic school (example: black/Mexican-American/Anglo) 36 parents will be interviewed (unless the fourth grade contains less than 12 students of one ethnic group).

In the 80 schools which are used in the student panel study (Sec. 8) a larger sample of parents will be selected instead of 12 parents per ethnic group. Parents of kindergarten, sixth and ninth grade students will also be surveyed in these schools, but only the third-grade parents will be used in the community reaction survey.

We also propose that three officers from the local parent group(s) (PTA or others) be interviewed in each school.

The survey of parents is designed to find out the effects of different desegregation policies, strategies, or plan features on their racial perceptions, attitudes, and feelings; willingness to support the schools and the school board; willingness to support public institutions; and educational preferences.

The parent survey should use a combination of mail, telephone, and personal interviews in order to reduce costs. The total amount of information to be gained from parent interviews is quite large, but there is no need to gather these data from the entire sample; therefore the instruments will be constructed by developing the total instrument set in the form of modules and assembling them in various alternate forms, as follows:

The *family background* section contains the conventional variables used as controls in analysis of student achievement and aspirations. The student panel survey requires a slightly extended version, since the control for family background must be extremely accurate in order to obtain good estimates of school impact on students. For this reason this section is slightly longer on some of the panel-parents interviews.

Parent *perception of school quality and participation in school activities* is the key cluster of items: it can be used to measure the amount of parent criticism of the school and the level of political participation.

We also hypothesize that parent *attitudes toward other ethnic groups* will both influence and be influenced by the quality of ethnic relations in desegregated schools. Since we are also interested in the extent to which minority parents are able to hold favorable attitudes toward their own ethnic group and have a sense of cultural identity, we propose that different forms of this cluster of questions be used for majority and minority parents.

The school district desegregation planning and process should affect parent *attitudes toward desegregation, the school system as a whole, and government.* Since this set of items will be analyzed at the district level rather than the school level, it is not necessary that the full sample of parents answer these questions.

In assessing the response of citizens to school systems policies, we are interested in their *electoral behavior,* both in bond and tax referenda and in other elections. Again, since this analysis will be done at the district level, a partial sample of parents is sufficient.

Finally, we are interested in knowing a great deal about parental *values and goals for schools* in order to determine what underlying principles influence adult attitudes toward desegregation, and also in determining what *school policies* or practices parents advocate for their schools. We are especially interested in the responses of the officers of parent groups in each school. Since these two sections necessitate a great deal of very careful questioning, they will be included in the personal interviews, with each respondent being asked to discuss either values and goals in general or school policies specifically, but not both.

As part of the family background section of the questionnaire, respondents should be asked whether they themselves attended desegregated schools. If this is done, the earlier Commission on Civil Rights report of the long-term effects of school desegregation can be replicated and improved with this study (Civil Rights Commission 1967, Crain and Weisman 1972).

This lengthy battery of questions can be administered by using alternating forms, so that no respondent receives the full battery. In particular, parental values and policy preferences will not be analyzed on a school-by-school basis and therefore need be administered to only 1000 respondents.

When these data are gathered, it will be possible to analyze the impact of community factors on parent acceptance of desegregation. This general topic covers

a whole host of questions, ranging from how different types of desegregation plans (with more or less busing, for example) affect parent response to how parent opinion responds to the behavior of community leaders. In this analysis, the role of television and the newspapers is critical. We need to measure the power of media effects. (For example, what happens if the media choose to either emphasize or play down school violence).

Four basic hypotheses that underly the design should be tested in the analysis:

1. Parental reaction to desegregation is strongly affected by community leaders, politicans, and the media.
2. The behavior of community leaders, politicians, and the media is influenced by their communication with key school officials.
3. Parental reaction is also influenced by the opportunities parents have to learn about what desegregation will mean; a great deal of fear on the part of both minority and majority parents is fear of the unknown.
4. Therefore the two most important thing school officials can do is co-opt community leaders into supporting the plan and creating opportunities for parents and students to experience desegregation on a trial basis in advance of the actual implementation of the plan.

Step 4: The Second Screener Survey

The second screener survey in the third year will use methods similar to those of the first screener but will be limited to the 160 school districts. It serves two purposes; to locate newly desegregated districts, and to gather data on school tax and bond referenda and school board elections.

The data on desegregation can be gathered with instruments very similar to those used in the first screener. The electoral data will be collected by means of a mail questionnaire sent to local sources of school election data (usually boards of elections, city clerks, county clerks, etc.).

Such aggregate election data are useful for comparison with opinion survey data to determine biases in the survey and differences between what people say they do and what they actually do. (For example, 15 percent more people say they vote for the winning candidate in national elections than actually do.) Also, because they will probably represent more than 30 points in time for many cities, the data will be useful for lag analysis of the fluctuations and trends in voting behavior in response to school desegregation. Thus the data may be able to catch rapid changes more easily than the panel study of voting opinions.

The location of newly desegregating districts is necessary in order to draw the sample for the studies that follow. In addition, the desegregation data can tell us which districts were able to expand earlier desegregation efforts, and which elected to discontinue desegregation or allowed schools to resegregate.

The analysis of the election data will complete the story of citizen reaction, since it will add a critical dependent variable—the political reaction of the citizens—to the earlier analysis.

Step 5: Case Studies of Desegregation Districts

Once school districts in the process of desegregating have been identified, we recommend that case studies be done. The advantages of the case study method are well known; it permits the researcher to incorporate rich and unanticipated data, and to develop new hypotheses. The disadvantages are equally well known; there is

little control over the investigator's bias, and no opportunity to determine the generalizability of the findings. Case studies are also very expensive. We recommend that a set of approximately 20 *comparative case studies* be carried out by a research team. This would enable each case to be compared with others, and the use of a team of researchers, recruited so as to provide a variety of competing philosophies, would help to reduce the bias in the analysis.

The case studies should focus on identifying particular events—public statements, news stories, or events occurring at schools—which seemed to influence public opinion toward desegregation, or which seemed to affect the success of the implementation of the plan. After they identify such events, the observers should interview various actors and informants in an effort to reconstruct the chain of factors that caused each event. This process should lead to the identification of critical variables. Each team should then share its findings with the other teams after every field trip, and a common pool of variables should then be jointly developed to be measured in each case study. We recommend that each case study, done in this fashion, involve only 8 to 12 person-weeks of field work the first year, and only 4 to 8 weeks the second year. When the research is completed, each case will contain a narrative of key events plus a coding of selected variables. Examples of such variables might be

- Commitment of superintendent to desegregation
- Number of staff planning meetings attended by principals
- Presence of support from citizens groups from integrated or changing neighborhoods.

The coding of these variables should be checked for reliability by having other researchers code the variables "blind" on the basis of the narratives.

The analysis of the case study then proceeds by combining the coded variables, the narratives, and the data from the earlier district leadership and parent surveys to produce a series of hypotheses about desegregation plan features and steps in the desegregation process that seem to bring about either (a) a good desegregation plan—one which is workable and creates the necessary conditions for successful desegregation; (b) public support for desegregation from both the majority and minority communities and their leadership; or (c) school staff commitment to the plan.

As with any method, the success of the comparative case study method is problematical. There are rules which should be followed.

1. The case studies must be conducted with a well-defined conceptual scheme. For example, the statement above that staff and public support results from observable events is a key element in such a scheme.
2. The team members should be highly intelligent, although not necessarily highly trained, and have good interpersonal skills.
3. The team members must be in close communication with each other and with other teams.
4. The analysis must be written with a conscious commitment to forcing the data toward policy-relevant findings. Descriptive cases are of very limited value.

The case study provides an opportunity not only to extend the earlier survey of district leadership but also to validate the survey. For example, in the course of testing hypotheses about district leadership, the case study team will be required to describe the type of leadership in the district; since this is a variable measured in the district leadership survey, a test of the validity of the survey can be performed.

Step 6: Third Screener Survey

The third screener survey in the fifth year brings the election data and school desegregation data up to date; the method is identical to that of the second screener.

Step 7: The Second District Leadership Survey

The second district leadership survey is largely a repeat of the earlier survey, limited only to districts which significantly desegregated in the intervening period. This should include all of the case study sites, and hopefully enough additional districts to provide an independent test of the case study findings. (In Fig. 6.1, 40 sites are suggested.) The case study hypotheses should be used to revise the instruments, and perhaps to alter the respondents used; but it is important to maintain enough common material to permit a clear before-and-after comparison.

Step 8: The Second Parent Survey

The second parent survey in the fifth year is limited to the same districts as the second district leadership survey, plus a set of schools which desegregated just prior to the first parent survey. (This group of schools is included because of a special longitudinal study of newly desegregated schools, described in Sec. 9.) This parent survey is limited to evaluations of the school their child attends, and to their overall attitudes toward school desegregation. Again, sample sizes cannot be determined, but this sample might be as large as 4000 parents if all schools in the 40 districts which desegregated after the first parent survey are included, plus 30 to 50 elementary schools which desegregated just prior to the beginning of the research (see Sec. 9), and if 12 parents of each ethnic group are selected from every third grade.

This survey permits a number of interesting comparisons. First, we have several before-and-after comparisons of parents in desegregating districts: we can compare parents in schools which have just desegregated with parents of similar social background in the same communities whose children were in segregated schools in the first survey; we can also compare them with parents whose children remain in segregated schools and parents whose children attend schools which have been desegregated for some time. Finally, we can compare all these parents with parents whose schools, which were newly desegregated at the time of the first parent survey, have now been desegregated 5 to 7 years.

The possibilities are rich, but this entire analysis must be done carefully in order to insure that the selection of groups for comparison does not introduce unnecessary bias.

The third screener survey, the second district leader survey, and the second parent survey together provide the data for a thorough analysis of the process of desegregation at the community level.

The eight steps of the data collection are dovetailed with the other studies proposed in the next three sections, as follows:

The first screener survey is used to select the sample of schools to be used in all other studies.

The district leadership survey provides data which can be used in the analysis of the large-scale longitudinal survey of schools (Sec. 7).

The parent survey is used in the large-scale survey, and also in the student panel study (Sec. 8). The case studies will to some extent overlap the proposed case studies of newly desegregated schools (Sec. 9). The second parent survey is combined with the parent survey in newly desegregated schools.

The community reaction survey is modular, and various components can be enlarged or eliminated as desired. However, these decisions will have implications for the other studies. Conversely, if decisions are made to alter or eliminate any of those studies, this may affect decisions about the utility of various components of the community reaction study.

In conclusion, the community reaction study can be thought of as a funnel: beginning with a minimal amount of data in 750 school districts, it finishes with a very intensive analysis in perhaps as few as 40 districts. The study combines four types of data: (1) data from districts and federal documents, (2) systematic interviews with community leaders, (3) parent surveys, and (4) case study interviews and observations.

The study is focused on a single question: What can be done to help a community and its school administration accept desegregation and implement it wisely and conscientiously? But these are only preconditions for successful desegregation; the success of desegregation ultimately depends on what happens inside the schools, and the next three sections deal with this.

7. THE LARGE-SCALE LONGITUDINAL SCHOOL SURVEY

The main data collection effort during the first two years of the project is a longitudinal study of over 500 elementary schools and the junior high schools and high schools which they feed. School data will be collected in two waves, one year apart: the first in the spring of the first school year, the second the following spring. Data from parents will be collected between the two waves.

A quasi-experimental design will be used, permitting us to compare matched sets of students to each other across pairs of desegregated schools and also between desegregated schools and segregated schools. In each school, data will be collected from principals and a sample of teachers and students of each ethnic group present. The basic data collection method is an informant-respondent questionnaire, in which the subjects will either respond both as informants, describing the school learning environment, or as respondents, supplying data about themselves. In addition, students will take achievement tests, and the research staff will observe the school.

Statistical analysis will be used to measure the impact of a large number of school factors on change in attitudes and behavior of both students and teachers between the pretest and posttest. As such, it will serve two purposes in the research program:

1. It will measure the impact of racial composition and the desegregation process on change in teachers and students.
2. It will measure the impact various school resources and school process factors have on changes in students and teachers.

These analyses will result in two major reports, to be completed three years after the program begins. In addition, this survey will serve two longer-range goals:

3. It will be used to select the schools for intensive study during the third to sixth years of the project.
4. It will be used to locate school resources or processes which seem effective in achieving integrated quality education, permitting us to study these factors more intensively during years three to six.

RATIONALE FOR A LARGE-SCALE, ONE-YEAR STUDY

As has been noted repeatedly, the research design must take into consideration the large variety of hypotheses to be tested, the advantages that different methods of data collection have over each other, and the very wide audience to whom the research is directed.

A large-scale study meets a large part of this need; it provides data representative of the experience of the entire nation, with relatively low sampling errors. We expect different regions of the country to have very different experiences with school desegregation; within regions we expect communities to differ; within communities we expect different ethnic groups to react differently. A basic sample of 550 schools clustered in 160 communities seems the minimum necessary.[1]

[1] The 550 schools will yield subsamples of schools (e.g., Southern desegregated schools, or Northern all-black schools) which will frequently have less than 100 cases. Since 100 cases required a correlation of 0.22 for significance (0.05, two-tailed), the sample size is not unnecessarily large for analysis.

Since desegregated schools differ in so many ways, a large-scale sample is necessary to locate sufficient numbers of desegregated schools of different types. One eventual goal is to produce carefully matched sets of schools which provide tests of a number of desegregation factors: busing versus alternative plans, desegregation beginning in different grade levels, staff ethnic composition, curriculum changes initiated, etc. A large-scale survey provides the screening with which to locate these well-matched test cases.

This study is large scale not only in the number of schools studied but in the number of concepts dealt with. The entire range of variables and hypotheses described in Sec. 1 of this report will be examined. With the present minimal state of knowledge about school desegregation, it is necessary to examine as many aspects of the issue as possible. This search of all aspects of the school process is intended to locate those interventions which seem most promising, so that they may be pursued in more intensive studies during the second part of the project. During this later intensive phase, new hypotheses will be developed based on data from smaller samples of schools. At that time, the analyst can return to this larger first study to validate findings with the relevant variables existing in the larger study.

The combination of a large number of variables and a large number of schools is necessary in order to locate as many as possible potentially successful interventions. Some school characteristics with important policy interest may appear in no more than 5 percent of the nation's schools at this time. Five percent of 550 schools is only 27, and we believe this is the minimum number for adequate analysis, either in this study or in a later more intensive project.[2]

A one-year longitudinal study seems the best compromise between the need for longitudinal data and the need for early policy-relevant and research-relevant results from the research. The advantages of longitudinal data are well known. While obviously the longer the time duration of the longitudinal design the more opportunity for change to be observed, a one-year design offers some special advantages. First, the experience in elementary school may change considerably from one year to the next for a child. By using a one-year design we are able to control the environment in which the student is placed. One year has proved to be sufficiently long to permit measures of significant amounts of change on some measures (Walberg 1972; St. John 1971). The use of a one-year design means that preliminary results from the longitudinal study will become available approximately three years after the first funding. We believe these results will be relevant to policy-makers; but even if they were not, they are necessary for the refinement of the design of the balance of the study.

SAMPLING

The sample is described in detail in Sec. 10. For the purposes of this section, the following points are important:

- The basic sampling unit is the elementary school attendance area.
- Students are surveyed at the beginning and end of the fourth, seventh, tenth, and twelfth grades. In junior high schools and high schools, a sample is drawn of 35 students of each ethnic group living in the elementary school attendance area. In the elementary school, the entire grade level is surveyed. This permits comparison between grades, even though the upper-grade schools cover a larger

[2] A subgroup of 27 schools will be significantly higher in student outcomes ($p < 0.05$, two-tailed) if the student outcomes is 0.4 school-level standard deviations higher; for cognitive achievement, this means the students would need to score ½ grade levels higher than students of similar SES in the other schools.

area than a single elementary school. Twelve parents of fourth-graders of each ethnic group will also be sampled.

- The sample consists of 130 "clusters" of elementary school attendance areas. Each cluster consists of two desegregated schools (i.e., schools with Anglos and one other ethnic group present) plus one segregated school for each ethnic group present in the two desegregated schools. Thus a biethnic cluster (e.g., a Mexican-American/Anglo cluster) would contain four schools: two biethnic, one segregated Mexican-American, and one Anglo. A triethnic cluster (e.g., a black/Puerto Rican/Anglo cluster) would contain five schools. The schools in the cluster are matched by using the census tract data on each ethnic group for the school attendance area. For example, the black students in the segregated school and the two desegregated schools will be of similar socioeconomic status (and, especially, similar in parent's educational attainment). This procedure is generally reliable, despite the fact that census tracts and school attendance boundaries do not match; we expect that when actual student social status data are gathered, no more than 10 percent of the clusters will be found to contain mismatched schools, and that these mismatches will result from the researcher's inability to correctly locate census tracts for students who are being bused.
- In some cases, school clusters will lie entirely within a single school district; in other cases they will lie in two similar and neighboring districts.
- We estimate that the 130 clusters will yield approximately 548 elementary attendance areas, of which 264 will be segregated: 130 Anglo, 100 black, 14 Puerto Rican, and 20 Mexican-American. There will also be 24 schools containing two minority groups: 12 with blacks and Mexican-Americans, 12 with blacks and Puerto Ricans. The remaining 260 desegregated schools will have Anglos mixed with blacks only in 170, Puerto Ricans only in 10, Mexican-Americans only in 24, both blacks and Puerto Ricans in 30, and both blacks and Mexican-Americans in 26.
- Eighty of the 548 attendance areas will be selected for the student panel survey (Sec. 8) after the first wave of data has been analyzed. In these schools, the parent sample and student sample will be enlarged.

This sample permits an analysis of the effects of desegregation with a quasi-experimental design. A full experimental design is politically impossible (we cannot randomly desegregate students) and a before-and-after design (surveying students before a school is desegregated and again after desegregation) would require a very long period of study, since some effects would not appear until students have experienced ten years of desegregation. The matched-pairs design is the strongest feasible design for assessing long-term effects.

The same quasi-experimental logic applies to other school characteristics as well. In particular, the presence of two desegregated schools in each cluster permits a matched-pair analysis of various techniques for improving the quality of desegregated schools.

MULTIPLE-INFORMANT DESIGN

The large-scale longitudinal study is limited by its size to relatively economical data collection methods such as questionnaires. This does not mean a great sacrifice in either the quality of data or the richness and complexity of the analysis. Modern survey methods in a longitudinal design can go far beyond the simple input-output analysis of *Equality of Educational Opportunity;* it can measure the dynamics of the

school environment and provide the data to permit analysis of nearly all the models described in Vol. I of this report. The two most powerful techniques are the pooling of individual student or teacher data so as to produce a single measurement for the whole school, and the use of informant as well as respondent questions.

Since we are concerned with school policy, our true unit of analysis is not the student but the school. Of course, we cannot "interview" our institution; we must rely on the human beings who are part of the school to answer our questions.

Since we are concerned with either the classroom or the school, data on individual students are of little interest to us. Therefore, the appropriate technique is to pool the responses of students to produce a single mean for all the students of a particular ethnic group in one classroom or in one grade in the school. Doing this enables us to minimize several types of errors, and to concentrate on that portion of the data which is relevant to social policy.

Consider an example: Suppose that we wish to test the rather simple hypothesis that the presence of prejudiced teachers is one of the major ways in which a school may lower the self-esteem of minority students.[3] To do so, we must first measure self-esteem.

But the response of any one student to questions about self-esteem may be subject to considerable error; he or she may misunderstand the question, or not want to answer it candidly, or perhaps on this particular day is ill and gives an answer he or she would not have given at any other time. If we compute a mean across all the minority students in the class, these individual errors, some of which produce inaccurately high measures and some of which produce unfairly low scores, will partly cancel each other out, so that the classroom-level result is more accurately measured. At the same time, the classroom-level result is more policy-relevant, since it ignores the wide variation in self-esteem resulting from differences in family background or individual personality, which is fundamentally uninteresting in any analysis of school policy. It is the difference between classrooms which is potentially policy manipulatable. An analysis of the aggregate level does not reduce our statistical power. It reduces the amount of variance in the data, but only by removing the variance which cannot be attributed to classroom effects; and it does not alter the degrees of freedom, which is the number of classrooms, not the number of students.

We can carry this reasoning a step further and conclude that a difference, not between classes, but between entire schools is what we should focus on (to do so avoids the severe problems created by the systematic assignment of students into ability-grouped classes). In an analysis at the classroom level, if the researcher finds that the self-esteem in one class is higher than in another after controlling on social class, he still does not know whether he has isolated a classroom effect, for if classes are ability grouped, the students may be in the first classroom in part because they have higher self-esteem. Unless students voluntarily choose their schools so as to create artificial differences (and this situation should be rare), it should be possible to compare schools by simply controlling SES, ethnicity, urbanism, and region. This provides the opportunity to aggregate the independent variable, teacher prejudice, as well; this removes the random individual-level difference and errors and permits us to concentrate on the collective racial attitude of the school. But this raises a new methodological problem, for if prejudiced teachers *systematically* bias their responses, the school-level data will still seriously misrepresent reality. For this reason, it is important to validate these responses by inviting all the actors in the

[3] To our knowledge, there is not in the literature a test of this rather obvious hypothesis. The data for a limited test exist in the *Equality of Educational Opportunity* study, but the analysis was never done, and the data may now be too out of date.

system (in this case, majority and minority students, teachers, and the principal) to estimate for us the level of staff prejudice in the school.

Questions can be divided into two categories: those which ask the subject to describe his/her own feelings or behavior (respondent questions), and those in which the subject is asked to describe the feelings and behavior of other people and the behavior of the institution (informant questions).

The gain in data quality when we move from respondent questions to informant questions is considerable. Of course, one cannot ask one person to report on the innermost feelings of another; but in viewing the school as an operating system, we are generally much more interested in the behavior of actors than in their true attitudes. Thus, the perception of a teacher's racial attitude by her students is in many ways more interesting than the teacher's true feelings on the matter.

When several informants are used to report on the behavior of any one actor, it is important that informants with different role-relationships to the actor be used. For example, if teachers report that their students are rowdy, we would like a second opinion from another perspective as to whether the students are genuinely rowdy or whether the teachers merely perceive them to be so.

When a number of different informants are used to report on their school, data of very high reliability can be gathered. For example, Ruth Narot constructed a scale of principals' attitudes toward desegregation in 200 high schools, using interview data from the principal and questionnaire data from the teachers and from black and white students, in a design similar to the one proposed here (NORC 1973). A total of six items were used in constructing the scale shown in Table 7.1, and the intercorrelations were sufficiently high to produce an overall school level coefficient of reliability of $+0.78$.

Having measured both variables, we are left with one remaining problem; if we find a positive relationship between staff prejudice and student self-esteem, can we argue that the first is a true cause of the second? We might establish reasonably high levels of confidence that one causes the other if we combine controls on spurious other factors (such as student socioeconomic status) with a longitudinal design which permits us to measure staff attitudes first, preferably before the students we are studying enter the school. For this reason, we have selected grades 6 and 9 for the pretest, since these are frequent school change points.

PRETEST SAMPLING AND DATA COLLECTION

The students to be interviewed during the spring of the first year will be students in the third, sixth, ninth, and eleventh grades. Thus, these data will be a pretest with which to measure change which occurs in the following year.

The data to be collected will consist of achievement tests and questionnaires administered to all elementary school students in that grade, and samples of 35 secondary school students of each ethnic group at each grade level; questionnaires given to eight teachers of each ethnic group from each school; an interview with the principal; and telephone interviews with twelve minority and twelve white parents of sampled third-grade children (see Fig. 7.1).

In junior high schools and high schools students are scattered among a large number of teachers. In order to enable us to carry out an analysis at the classroom level, the students should be identified by their English classrooms and their English teachers interviewed. This will enable us to use English classrooms as a special target in the analysis of school and classroom characteristics. Schools will be contacted in advance to obtain lists of students with addresses, in the grade levels to

Table 7.1

ITEMS IN PRINCIPAL DESEGREGATION ATTITUDES SCALE

Item	Principal			Teachers	Black Students	White Students
	General Racial Attitude	Desegregation Helps Whites	Favors Desegregation	Principal's Attitude	Principal's Attitude	Principal's Attitude
Principal: general racial attitude[a]	X	0.45	0.50	0.47	0.33	0.35
Principal: desegregation helps whites[b]		X	0.51	0.38	0.27	0.30
Principal: favors desegregation[c]			X	0.45	0.21	0.32
Teachers: principal's attitude[d]				X	0.34	0.43
Black students: principal's attitude[e]					X	0.30
White students: principal's attitude[f]						X

SOURCE: Unpublished data from the Evaluation of the Emergency School Assistance Program; see NORC (1973) for study description.

NOTE: Reliability score = 0.78; average correlation = 0.37. All correlations are product-moment, from data aggregated at the school level.

[a] A scale rating the principal's personal views of integration including whether or not he would live in an integrated neighborhood, his feelings on interracial marriage; the amount of prejudice in this country, and the good or harmful effects of the civil rights movement (+ = more liberal).

[b] Principal's response to whether whites do better in integrated schools (+ = whites do better).

[c] Principal's feeling about desegregation (+ = likes it).

[d] Mean of teachers' evaluation of how principal feels about desegregation (+ = likes it).

[e] Mean of black students' evaluation of how principal feels about desegregation (+ = likes it).

[f] Mean of white students' evaluation of how principal feels about desegregation (+ = likes it).

be studied. The school staff will be asked to identify the students by ethnicity. This list will be randomly sampled to select 35 (if possible and allowing for absenteeism) students of each ethnic group being studied, who live within the attendance area of the elementary school being surveyed, thus making the secondary and elementary students in the survey comparable. The contractor may prefer to survey all students; in this case the instruments may be administered in regular English classrooms. If the students are sampled, they should be relocated to a common meeting room—the library or auditorium—for group administration. All questionnaire administration and testing will be done by the research organization's staff; no school staff should be in the room.

Every school is different; the local research staff should be encouraged to adapt procedures to the desires of the principals and the convenience of the school. However, this does not include the actual mechanics of test and questionnaire administration, which must remain standardized.

The distinct linguistic and cultural background of Mexican-American and Puerto Rican children poses special problems to the data collection effort. Although both groups speak dialects of Spanish, they come from widely divergent cultural heritages. It seems that Mexican-Americans, as a group, perceive the value of desegregation differently than do Puerto Ricans (as noted in Section 5 of this report). The former see desegregation as a means of improving the quality of education of their children, whereas the latter feel that desegregation of schools is only secondary to the need of quality education. Nevertheless, both groups value highly the implemen-

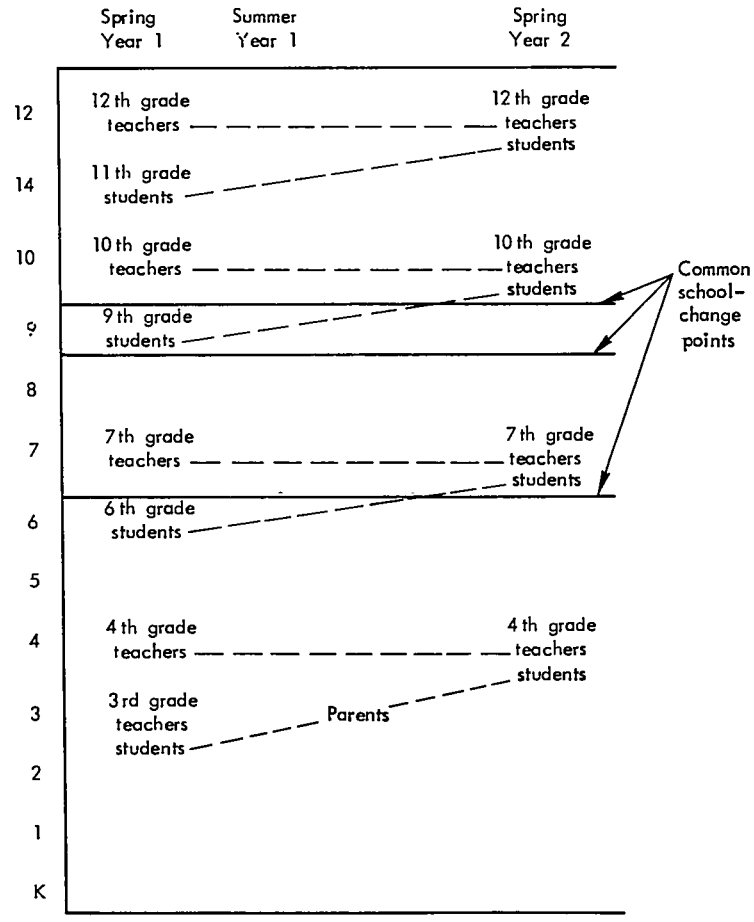


Fig. 7.1—Respondents in the large-scale longitudinal school survey

tation of bilingual/bicultural programs as a means of providing quality education for their children.

To insure the reliability and validity of the data to be collected it is imperative that the contractor responsible for data collection be sensitive to the language and culture of the Mexican-American and Puerto Rican minority groups. More explicitly, the contractor should not only have a staff that speaks Spanish but also one that is fluent in the vernacular spoken by these two groups in various geographical regions of the country.

The procedure for selection and administration of instruments to students and teachers in schools with significant numbers of Spanish-language students should include the following: (a) instructions will be given both in English and in Spanish; (b) instructions will be given orally if necessary in the younger grades in the appropriate dialect of Spanish; and (c) instructions and questionnaire instruments will be printed in both Spanish and English on the instruments. The cognitive tests will be in English only, however.

Spanish translations of instructions and questionnaires are not provided in this report. The translations of questionnaire terms should be literal, avoiding colloquial

phrasing. Students should be asked which language they used in filling out the questionnaire, and the analysis should include an analysis of response differences due to language used. We have designed a short instrument given in Appendix C to measure English-language proficiency for nonnative speakers of English.

Students must give their names to the research team to permit the establishment of a "clean" panel by removing those students who do not reappear in the sample for the posttest. A recommended procedure is described in Sec. 8.

The teachers presently teaching the sampled students and those likely to teach the students the following year should be interviewed. This should be done so as to maximize the number of teachers who are interviewed at two points in time. In some cases this may necessitate interviewing more than eight teachers of one ethnic group; in other cases it will be impractical to interview every teacher who might teach a fourth-grade class or seventh-grade class the following year. In junior high schools and high schools a different procedure is necessary. If the schools are organized in a traditional 6-3-3 structure, then secondary school teachers should be sampled in order to maximize the chances of interviewing the teachers who will have responsibility for basic classes (especially English, but also math or social studies) for tenth-grade and twelfth-grade students. In junior high schools which serve only the eighth and ninth grades it is not necessary to interview teachers and principals because no longitudinal analysis of the educational process in these schools will be possible.

The teachers should be surveyed with a written questionnaire; the principal by face-to-face interview.

Parents of surveyed students are to be sampled randomly from the student lists, and letters of introduction mailed to them before they are interviewed.

In the course of contacting the school, interviewing the principal, drawing the sample, and administering the questionnaires, the research team will have an opportunity to observe the principal's personal style and the overall climate of the school. The team's perception is quite valuable, since it represents the view of an unbiased outsider to the system. For this reason, the team should be instructed to systematically observe a number of parts of the schools (particularly student and staff behavior on the playground, outside the school, while boarding buses, and in the cafeteria, halls, principal's and teachers' lounge) and report this observation on a standardized questionnaire form. Each member of the research team will fill out the questionnaires, providing the data to compute reliability coefficients for these data. We anticipate that many of the items will have reliabilities high enough for analysis. The school physical environment may be very important (e.g., pleasantness, openness, color, cleanliness, privacy, absence of crowding, good acoustics). The uses of physical space and other aspects of the physical environment have potentially important psychological impact, and the study should take note of these.

POSTTEST SAMPLING AND DATA COLLECTION

Data collection for the posttest is similar to that for the pretest.

There is an additional sampling problem created by the fact that in secondary schools the students being surveyed are from one attendance area and therefore are not representative of the whole school environment. This problem is unavoidable when students from several elementary schools (including the one we are studying) come together in a larger junior high school, as would frequently happen between sixth and seventh grade.

Our interest is in the longitudinal analysis of the particular group of students who were pretested—thus an analysis of a group of elementary school students as they are influenced by their first year of junior high school. However, we must have a representative sample of the students in the posttest school in order to measure accurately the student environment in which our panel of students has been placed. To do this a supplemental sample must be drawn for the posttest, consisting of thirty students selected in the following fashion:

In the seventh and tenth grades, the English class of each student in the panel is determined, and a nonpanel student sampled from that English class for each student in the panel. This generates a first-stage sample of exactly the same size as the number of survivors of our original panel. This first-stage sample is then randomly sampled so as to reduce it to thirty students. The resulting sample will have the same distribution across English classes as the panel and therefore is the best estimate of the student environment affecting the English class experience of our students. The analysis procedure for these types of data is described in a Appendix A.

At the time of the posttest elementary school teachers can be asked questions about specific students and their classes, thereby providing measures of teacher attitudes toward students of different ethnicity and achievement. The exact procedure is given with the questionnaires for the survey in Appendix C.

DATA TO BE COLLECTED

With this procedure it is possible to generate a rich store of data on student outcomes on the independent variables of school personnel and structure, and on the intervening variables of school process with which to explain student outcomes. Instruments for the first wave of the study are given in Appendix C. They are to be field-tested and revised before administration.

The first-wave student data has two main purposes: to gather base-line data on student outcomes and base-line data on school processes. The basic student outcomes are:[4]

Measures of student learning
Measures of sense of self-esteem and self-acceptance
Sense of trust in and control of environment, expectations of others
Altruism and cooperative behavior
Anxiety and happiness
Aspirations and achievement orientation
Attitudes toward school, interest in learning
Attitudes toward other ethnic groups
Attitudes toward own ethnic group
Occupational and academic aspirations
Civic and political knowledge and sophistication
English-language proficiency
Motivation to learn English and Spanish
Students attitudes toward the learning of English and Spanish

[4] The instruments now designed do not adequately measure a few of the concepts listed here; hopefully additional items will be developed by the data collection contractor.

Measures of school process, to be collected wherever possible simultaneously from observers, teachers and students, include:

- Interracial contact among students
- Amount of school work done by students, including homework
- Quality of adult-student interaction in the school, emphasizing rewards and punishment students receive
- Amount of enjoyment students receive from school work
- Extent of parental involvement
- Measures of racial tension
- Teacher and principal's attitudes towards their students learning of English and Spanish
- In high schools, student dating behavior including interracial dating
- Measures of teacher behavior and teacher attitudes toward students
- Other process variables as identified in Sec. 4

The independent variables can be measured from the teacher and principal questionnaires. These include:

- Teacher and principal background
- Principal administrative behavior
- Degree of ability grouping
- Type of classroom organization and degree of individualization of instruction
- Extent of extracurricular activities
- Extent and type of multiethnic programs
- Extent of remedial programs
- Extent of parental involvement
- History of the desegregation plan for this school
- Political decisionmaking in a school regarding policies and school goals, with emphasis upon extent of student and parental involvement
- Degree of administrative decentralization in the district
- Goals set for the local school by the school district administration
- Goals set by the principal for his school
- The existence of bilingual (English as a second language)/ bicultural programs in the schools
- Other variables defined in Sec. 4

The second wave of the school survey will be similar to the first wave. Many measurements will be repeated in order to gain measures of change in both student outcomes, teacher behavior, and school process.

The construction of the second-wave instrument can also be guided by preliminary analysis of the first-wave data. The second wave will emphasize teacher behavioral data, since it is only during the second wave that we can guarantee that we have interviewed the relevant teachers for each student. Additional process variables will also be added, since now both teachers and students can report upon the past school year, which should have the greatest impact upon student outcomes. Student family background variables and student reports of earlier school experiences can of course be dropped from the second wave. The research team should be

given considerable freedom in developing the second-wave instrument in order to take advantage of both their expertise and the knowledge gained from preliminary analysis of the first-wave data.

DATA ANALYSIS

The first task of the research team will be to analyze the first-wave data to verify that the matched sets are reasonably accurate in their matching. Secondly, there should be a preliminary search using standard multiple regression techniques to locate promising school independent variables and process variables which seem to influence student outcomes. While this analysis lacks the power of a longitudinal design, it can tell us a great deal about schools, and this information can aid us greatly in deciding what aspects of the school to emphasize in the second-wave data collection. (It should be recognized that even if the first wave of the study were all the data to be collected, we would have a data set richer than any we have had in the past.)

Finally, the first-wave data must be analyzed in order to determine which schools should be selected for inclusion in the various subsamples for the more intensive studies.

After the second-wave data have been collected, a standard longitudinal analysis of school outcomes can be undertaken by using a wide variety of statistical and analytic techniques. A careful study of the problems of longitudinal analysis (the gain score literature and the work of Donald Campbell (1972a)) convinces us that there is no single obvious solution to the problems of bias, regression effects, etc.

There are two problems here. The first is the gain analysis problem: basically, there is no technique which will permit a good comparison of gains between two populations if their starting points are noticeably different. Gain scores (the difference between the posttest and pretest) will generally favor the group which starts highest, since this group will frequently have a steeper rate of growth; thus if a remedial reading program is administered to low achieving students, the group may still gain less than a higher-achieving control group, leading us to conclude that the remedial program has failed. If we attempt to match students who receive special programs with low-achieving students from the control population, we are likely to be matching remedial students who scored unusually high with control students who scored unusually low; on the posttest, each group will drift toward their normal scores, and again the program will show a negative effect. This is "regression toward the mean." These problems cannot be avoided by any statistical techniques, although it may be possible to develop techniques which are less sensitive to these problems. We recommend that the research team investigate the relative merits of covariance analysis, standardization, and superstandardization to determine which is most effective with these data. However, the most satisfactory defense against these problems lies in the reliability of the pretest measures and in the use of quasi-experimental and experimental designs. We have therefore recommended that (1) data be aggregated to the school and classroom level, which increases reliability; (2) segregated and desegregated schools be matched on the basis of census data, rather than pretest scores; (3) if possible, potentially effective treatments for desegregated schools should be randomly assigned in a full experimental design, as recommended in Sec. 12.

There is a second statistical problem which is perhaps more important: whatever statistical procedure is used yields a measure of effect—of an innovation, of desegregation, of school itself. Unfortunately, none of these measures unambiguous-

ly tells us whether an effect is large or small. *Equality of Educational Opportunity* used the unique percent of variance explained as a measure of effect. There are two problems with this. First, the measure is conservative in dismissing all shared variance. More importantly, the measure encourages the reader to assume effects are small. The statement "only 20 percent of the variation in achievement is attributable to school effects" sounds as if it has a common-language meaning; in fact it is a statistical statement with no verbal interpretation. Crain has pointed this out (NORC 1973) by presenting graphical presentations of school effects which do not seem small (although the percent of variance explained remains the same, of course). Coleman (1973) has also addressed this problem, presenting common-language questions about school effects and attempting to construct statistical measures appropriate to them. This is the reverse of the usual procedure, which is to take a statistic and fit a common-language interpretation to it. Coleman's approach seems to argue that school effects are stronger than he himself felt them to be. We believe that Coleman's work is important and should be pursued by others. If no new statistic is developed, we believe that the statistic most likely to be interpreted wisely is the regression slope coefficient. The slope presents directly the amount of change in the outcome variable attributable to a certain amount of change in the independent variable. If the units of measurement for the independent variable are properly chosen, the regression coefficient should be interpretable. We believe the percent of variance explained is in the wrong dimensional units. For analysis of variance, measures paralleling the regression coefficient in the dimensionality should be used.

Statistical problems are not the most difficult part of the analysis. One cannot always interpret responses in terms of their face-valid meanings, and this is probably the most difficult technical problem in the analysis. For example the first evaluation of the Emergency School Assistance Program found that schools with higher levels of teacher in-service education programs had teachers who were generally more pessimistic about the progress that their school had made toward good race relations (Resource Management Corp. 1973). The second evaluation of ESAP found that more black high school students took a more anti-integration stance in schools which by some measures appeared more egalitarian in their race relations (National Opinion Research Center 1973). In both cases there is at least a possibility that the responses reflect changes in the ways in which students and teachers perceived their environment rather than true changes in attitude. The well-trained teacher may take a more analytical view of his environment and may set higher standards for success. Students may react to racial equality by enjoying the freedom to express their previously latent hostilities toward other groups. In both cases the implication —that the particular program had negative effects—is too simple. In general, these problems can be dealt with by using reports from other informants as to the perceived behavior of various groups.

The analysis is multipurpose: it is concerned both with analyzing school process to give us a better basic understanding of what happens in desegregated schools and how this contrasts to segregated schools; it is also concerned with locating those points at which federal and other extraschool policy-makers can intervene to influence school quality. This should not be taken as implying that we are interested only in obvious forms of intervention such as the awarding of Title I grants to a school. Extraschool pressures from the civil rights movement, from the school board itself, and from federal requirements specifying the need for citizen involvement are all legitimate forms of intervention which may in the long run have as much effect as any Federal aid program, maybe more.

PROBLEMS OF SCHOOL, STUDENT, AND TEACHER ATTRITION

We can expect the combination of mobility and absenteeism to reduce the posttest sample of students by 10 to 50 percent, and the teacher sample somewhat less. In the posttest, no effort should be made to replace the sample losses of students, but additional teachers should be surveyed to maintain this sample size.

The problem of loss of schools from the sample through noncooperation is very serious, and one for which we have no solution. There is little that the research institute can do beyond applying a high level of tact in its negotiations with the school systems. The decisions which will influence cooperation will be made by the client agency, not the research staff, and are discussed in Sec. 15.

The magnitude of the problem is highlighted by the difficulties of *Equality of Educational Opportunity.* In that study the sample was sharply biased through the loss of many school districts, including a number of very large districts such as Chicago and Los Angeles.

Analysis of samples biased by attrition can be done only by gathering data on the missing districts. In some cases, this is done by selecting a subsample of the losses and pursuing them with more expensive methods. There do exist some data from the census and from the leadership and screener survey which do not require district permission. These data can be used to assess the nature of the bias, and this information can then be used to weight the present cases to cause the weighted sample to more closely resemble the nation. Secondly, the research team and client agency can draw a subsample of noncooperating districts and bring to bear a variety of inducements to change their mind. If a high cooperation rate is obtained from this subsample, the research team would have a good sample to use to assess the bias created by noncooperation. A less expensive procedure which is commonly used is to divide the cooperating districts by degree of cooperation (separating those which initially agreed to cooperate from those which debated the issue before agreeing, for example). If, for example, we find strong effects of certain programs in the highly cooperative districts and weaker effects in the reluctant ones, this would lead us to be cautious in generalizing the effect to noncooperating districts.

It should be borne in mind that we are not mainly concerned with estimating national norms; we are concerned rather with understanding the processes of school desegregation, and it is at least possible that the processes may be sufficiently similar in the lost districts to permit generalization, even though the lost districts are not typical of the nation. (For example, they may be more segregated, but segregated for the same reasons as the segregated cities where data were obtained.) Therefore, in order to investigate this, one must not attempt to examine differences in means between refusing and nonrefusing cities, but differences in measures of association. This can be attempted in two ways: by analyzing the leadership sample correlations in the two populations, and by comparing correlational data for schools when the data are and are not weighted to account for nonresponse.

CONCLUSION

The large-scale longitudinal survey of schools is intended to provide at least preliminary data in answer to a wide range of questions:

- What are the usual effects of desegretation?
- How does one identify a school which has been successfully transformed from desegregated to integrated?

- What policy instruments seem successful in making desegregation work?
- If desegregation is successful, what are the likely consequences for children?
- If desegregation is unsuccessful, what are the consequences?
- What schools have useful properties which make them worthy of inclusion in any of the various follow-up studies?

The study should be very valuable, but it has the following drawbacks:

- The study period is only one school year; some writers have argued that this is too short a time to observe meaningful change.
- Because the study is so large, it cannot gather enough data on any one topic; it will only scratch the surface in a number of areas.
- Because the study is so broad, and uses the survey method exclusively, it will not provide good enough data to describe the climate of the school, the nuances of day-to-day behavior which we have called the process dimension of the school.

The substudies in the following two sections are designed to overcome these weaknesses.

8. THE STUDENT PANEL STUDY

INTRODUCTION

The longitudinal survey described in Sec. 7 provides baseline and one year change data for a large representative sample of desegregated schools and non-desegregated schools selected for comparison purposes. As such, it will provide a basic picture of the state of school desegregation throughout the country. Equally important, the longitudinal survey will provide information that enables a selection of schools for substudies that can more efficiently investigate special issues with greater intensity and rigor.

The Student Panel Study (SPS) is one such substudy proposed to extend the scope and duration of the longitudinal survey. The SPS is designed to make the following contributions to the understanding of the school desegregation process:

- An assessment of relatively long-term as opposed to short-term changes in student behavior.
- An investigation of the extent to which parents play an intervening role between desegregation and student outcomes.
- Utilization of a quasi-experimental design that will enable more rigorous tests of cause-and-effect relationships suggested by the longitudinal survey.

All of these goals are addressed by the one-year large-scale study, but some writers have argued that one year is too short a period to assess student change. If the effects of the school are too slight, then the effects over one year may be too small to be unambiguously detected. This study is designed to assess change over at least a three-year period. It is also designed to interview a large sample of parents in order to measure more accurately the home background's influence on the student. Interviewing parents and following students for three years are both expensive; we have compensated by sharply reducing the sample size. Thus the long-term, small sample of the SPS and the short-term, large sample of the large-scale study complement each other.

This section will first specify these goals in more detail and, second, spell out a research design to attain them.

GOALS OF THE STUDENT PANEL STUDY (SPS)

The goals of the SPS can be grouped under three broad categories of policy-relevant issues. First, the SPS aims to discover the *effects of desegregation* on changes in student behavior over at least a 3-year period. The basic policy question here is the degree of student change in desegregated schools as compared to student change in nondesegregated schools. Second, since we cannot assume that *all* desegregation policies will result in successful change, the SPS will investigate *models of desegregation* in order to discover intervening or mediating conditions that explain changes in student behavior. The policy issue here concerns the determinants of successful desegregation policies. Third, the SPS can be used to investigate the impact of desegregation relative to other types of instructional programs deemed highly desirable by some minority groups, such as bilingual/bicultural instruction.

This policy question concerns the value of desegregation as a school policy as contrasted to or in conjunction with other types of school intervention programs.

Effects of Desegregation

One of the basic policy questions that has been raised in the past several years is the extent to which desegregation leads to increased educational and social opportunities for minority groups. Most prior research, however, has adopted rather narrow definitions of opportunity, mostly in terms of academic achievement (Coleman 1966). While academic achievement is still an important outcome characteristic, we propose to use a broad definition of opportunity that will include other kinds of academic behavior, race relations, and psychological outcomes. In addition, the SPS plans to use frequent measurement over a sufficiently long time period to establish clear trends in those various student outcomes.

The basic student outcomes were presented in Sec. 7. Here we reveiw them, highlighting those parts most relevant to the SPS.

Academic Behavior. We define academic behavior as that cluster of student outcomes most directly related to the formal goals of primary and secondary schools; we take these goals to include training in basic cognitive skills and preparation for post-high school educational and vocational choices.

The basic academic outcomes that seem to us crucial for equal opportunity, therefore, include academic achievement in basic skills such as reading in English; motivation for learning those skills relevant to one's career choices; a high but realistic self-concept in relation to one's academic achievement and educational or vocational choices that are in accordance with one's interests and aptitudes. Among this latter set of outcomes the decision about college and the actual start of college are two high-priority criteria.

Race Relations. While the primary goals of schooling may be of an academic nature, the school is also a primary socializing institution where students develop values and behavioral standards that they will carry with them into adult life. Accordingly, a crucial policy question is the extent to which desegregation affects race and ethnic group relations. Proponents of desegregation have cited improved race relations as one rationale of school desegregation; opponents have claimed worsened race relations in arguing against the policy.

Given the many different interpretations different groups may have about this concept, including the possibility that different racial and ethnic groups will have different definitions of what constitutes good race relations, it is necessary to construct a broad set of criteria to assess race relations. We can discern and classify important outcomes according to *cognitive, affective,* and *behavioral* domains.

In the *cognitive* domain we are concerned with changes in both knowledge and beliefs about different groups as a result of desegregation. Psychological theory has long held that one group's lack of knowledge about another group is the primary basis of ethnic stereotyping and prejudice (Allport 1954). Whatever else may occur in the race relations realm, then, increased knowledge about ethnic groups other than one's own group would be viewed as a positive outcome of desegregation. Under the cognitive realm we would also include beliefs and opinions concerning the desirability of racial and ethnic balance in schools, communities, and later life situations.

Under the *affective* rubric we are concerned with attitudinal and emotional changes in a group's feelings toward another group. This would include levels of intergroup fears, anxiety, and tension that might change as a result of integration. In this regard we must realize the possibility that these affective states may worsen

during early stages of desegregation due to lack of knowledge and prior intergroup experience, but we would expect that a successful program would see them improve in the long run. This underlines the need for a relatively long-term panel study. In this category we would also investigate changes in prejudice and general intergroup attitudes and feelings including views towards desegregation.

By *behavioral* we mean those outcomes that have some basis in concrete actions apart from beliefs and feelings. The two behavioral outcomes of foremost importance are changes in intergroup contact and discrimination. Successful desegregation should lead to increased levels of volitional social and informal contact, or at least to stable (rather than declining) levels of such contact. Contact might include friendship choices, extracurricular activities, lunchtime or recess association, and schoolwork cooperation. We must stress that there is not a great deal of scientific evidence about changes in intergroup contact patterns as a result of school desegregation based on real-world situations (Pettigrew et al. 1973; Armor 1973); therefore the knowledge gained from the national desegregation study is likely to establish a set of propositions about contact not heretofore verified. For this reason we must be cautious about giving positive or negative interpretations to various patterns of change. There is some evidence, for example, that in many circumstances there is a tendency for groups to "resegregate" in their informal social relations at school (Gerard and Miller 1971; Armor 1972; Patchen and Davidson 1973); but in the short run this may not necessarily be viewed as a negative consequence. As far as ethnic discrimination is concerned, we are aware that this is a behavior that is very difficult to measure without an extensive observational methodology, and we may have to rely on self-reports or on perceptions of discriminatory behaviors.

Personality and Psychological Outcomes. There are a number of nonacademic outcomes other than race relations which have been cited as important in school desegregation. Some of these are viewed as potential influences on more explicit academic behaviors; others are viewed as important to a successful adjustment to adult life. Most prominent in this area might be locus of control (internal vs. external) or fatalistic outlooks. Persons who have attained socioeconomic success have been shown to have high levels of internal control and nonfatalistic outlooks, although it is not always clear which state came first. The longitudinal design we propose may help untangle this particular causal order. Since it will provide measures of various psychological factors and school success and aspiration factors at several points in time *permitting a cross-lagged analysis.* We are also concerned with such outcomes as general self-esteem (other than academic), happiness, and other general personality measures such as trust, sociability, etc. (Crain and Weisman 1972). Nonracial school factors affect these outcomes as well as racial factors; this means that racial and nonracial causal variables will need to be combined into a multivariate analysis.

Under this category we also want to place those behaviors relating to discipline and conformity to school norms. There is much discussion of so-called "discipline" problems in connection with school desegregation experiences. It is important, therefore, to carefully measure behavioral deviance to make sure that discipline problems are truly violations of important school rules as opposed to a violation of one group's particular social mores. Regarding true discipline problems, while it is unlikely that school desegregation is a social force of sufficient power to create an antisocial personality (or a pre-social one), it is possible that some policies exacerbate students with those tendencies and hence tend to increase discipline problems. Whatever the true state of affairs, there has been enough discussion of the issue to warrant a careful assessment of this outcome.

Models of Desegregation

The goals suggested in the previous subsection are concerned with the overall effects of desegregation. The national desegregation study proposes to go beyond this to describe models of desegregation which can differentiate determinants of successful or unsuccessful desegregation policies. Basically, the questions we ask here are concerned with intervening conditions between desegregation (as the primary independent condition) and student outcomes. That is, are there crucial conditions which must hold in order to obtain positive outcomes? While these intervening variables may not always be controllable by policymakers—and hence the need for an overall evaluation of desegregation—they are nonetheless important for a full understanding of the scientific process and for future policy directions.

We will organize intervening conditions and determinants of success under three main headings: *plan and environment, family, and school and teacher.* It should be stressed that the SPS itself will not necessarily measure all the variables within these domains (e.g., the main longitudinal survey will do some of this). But we do want to describe those conditions viewed as potentially important for influencing long-term student change, since this will affect the design and analysis strategies of the SPS.

Plan and Environment. The most important intervening condition—the more so because of the paucity of information about it—concerns the type of desegregation plan being implemented. The major variations in plan include court-ordered (including plans ordered by state agencies), community-initiated, and natural (i.e., due to residential patterns or density). The community-initiated plans further divide into voluntary where no students are required to be reassigned, and mandatory; it may also be useful to distinguish plans which are implemented only after intense pressure from those which seem more willingly adopted. It is quite possible that the success of desegregation depends partly on the way in which the desegregation process is initiated. One might speculate, for example, that white flight and poor race relations are more likely in a court-ordered plan (which presupposes community opposition) than in the case of natural desegregation that arises from long-standing residential patterns.

In addition, certain other extra-school factors—such as region, urban-rural, community attitudes, and socioeconomic and ethnic composition of the community —are likely to influence certain student outcomes. We group these environmental variables with plan variables because they are not easily separable in practice: court-ordered plans are more likely to be found in the South; and those few in the North are urban settings likely to be accompanied by substantial community opposition, which may explain why the court got involved in the first place. Community-initiated mandatory plans (e.g., Riverside, White Plains, Ann Arbor) are likely to be accompanied by substantial community support and a relatively small proportion of minority students. Voluntary plans frequently reflect either an ambivalent community (e.g., Boston) or a community with a substantial proportion of minority students (e.g., Los Angeles). Finally, natural desegregated (non-induced) schools—especially junior and senior high schools—are likely to be found either in the rural or small-town North (e.g., Sharon, Mass., or Victorville, Ca.).

Family. Family characteristics are likely to be important intervening conditions in two primary ways. First, the attitudes of parents regarding the desegregation plan and toward other ethnic groups could very well determine positive or negative outcomes in student race relations. Second, the socioeconomic level of the family and parental values about education[1] are likely to affect academic outcomes

[1] For example, a Mexican-American or Puerto Rican parent's view of the relative importance of school desegregation as opposed to or in conjunction with bilingual/bicultural education.

such as achievement, aspirations, and college attendance. Family SES may also affect student race relations, since it is often argued that lower and working class Anglo-Americans have more ethnic prejudice than middle-class Anglo-Americans.

In this latter connection, it should be stressed that an analysis of desegregation effects should not stop with an assessment of the effect magnitude; full policy impact cannot be determined without an assessment of the magnitudes of other outcome determinants. Family background characteristics are known to have substantial effects on academic behavior, and therefore are primary candidates for comparison criteria.

School and Teacher. The Coleman report (1966) found that the effects of school and teacher characteristics on academic outcomes were not as strong as family background effects and were not as strong as many educators have assumed. Nonetheless, there are at least two reasons why they should be examined for their potential mediating influence on desegregation outcomes. First, we are studying outcomes other than academic achievement; school and teacher effects may have more importance for noncognitive outcomes, particularly for race relations. Second, the Coleman report was not longitudinal and had certain other methodological defects; another attempt with a more rigorous design may turn up school and teacher effects not discernible in the Coleman study.

The school characteristics of interest to the SPS are no different than those measured in the main longitudinal survey, and no special data need be collected except that school resources should be measured at least one more time during the relatively long panel study period to measure changes in school resources. Teacher characteristics are of greater concern to the SPS, particularly those characteristics that might be related to noncognitive outcomes. While it would not be appropriate to conduct a full classroom process study in the SPS, several critical teacher characteristics should be given special attention: ethnicity; ability (e.g., verbal achievement in one and sometimes two languages; professional background and experience); attitudes about desegregation, other ethnic groups and the language used by other ethnic groups; rankings of panel students' ability, discipline problems, and popularity. All but the latter ranking variables will be assessed as part of the main study and in the additional years of the panel. The ranking of students according to dimensions of school success and adjustment are intended as measures of teacher expectancy, one of the most important intervening conditions cited in the current literature. These rankings can also be used in conjunction with objective information about the same dimensions to derive an indirect index of discrimination.

GENERAL RESEARCH DESIGN

The SPS is conceived as a quasi-experimental design that follows a panel of students and parents over time. The panels will be drawn from a small number of schools selected intentionally to provide adequate treatment-control comparisons. While the main longitudinal study selects quartets or quintets of schools matched on basic demographic and socioeconomic characteristics of school attendance zones, the intent of the SPS is to refine this by matching each desegregated school with other desegregated and segregated schools with respect to critical socioeconomic and academic characteristics based on data collected in the longitudinal survey. In this way it should be possible to obtain better control groups and hence to make more rigorous cause-and-effect inferences about the effects of desegregation (compared to segregation) or the effects of a particular type of desegregation (compared to another type).

The intent is to assess the student panel once a year for at least three years. Although the main longitudinal survey will follow students for one year, it is assumed that this is not long enough to establish definite trends and to make final policy judgments. Therefore, the SPS will make student assessments each year for at least three years. Parents will be assessed twice, once at the beginning of the panel study and once at or near the end. In this way it will be possible to use both initial parental characteristics as well as changes in parental characteristics as intervening variables in student outcomes. Also, changes in parental attitudes can be used as outcome variables as well. The first wave of assessment (except for parents) is provided by the first year of the main longitudinal survey.

School Selection

The main longitudinal survey will select quartets and quintets of schools matched for demographic and socioeconomic characteristics using U.S. census data for the school attendance zone. While this should produce more comparable schools than a simple random sampling of school districts, it is expected that comparison schools may still differ in crucial ways that hamper cause-and-effect inference. For example, the census data do not permit matching of students according to academic achievement, a characteristic of great importance in the assessment of academic outcomes. Also, in some cases the socioeconomic level of the students in a school may differ appreciably from the socioeconomic level of persons living in the attendance zone; this may be especially true in school districts where substantial numbers of students attend private schools, or where a busing plan is so complex that the research team cannot define the attendance zone for the school.

Given these conditions of matching in the main longitudinal study, and given that the SPS will follow a much smaller sample of schools, it makes sense to use data from the main longitudinal survey to refine the matching process. As in the main survey, we envisage sets of four or five schools (quartets or quintets) composed of two desegregated schools and two or three segregated schools. For a quartet, one segregated school would be predominantly white but matched to the white students in the two desegregated schools; the other segregated school would be predominantly minority and matched to the minority students in the two desegregated schools. The quintet is similar except that it envisages two minority groups in the desegregated schools (say, a black minority group and a Spanish-speaking minority group). Thus three segregated control schools are required, one matched to the white students and one each matched to the two minority groups.

Perhaps the two most important characteristics to match are socioeconomic level of the families sending children to the school and the average achievement level of the students in the school. It might also be desirable, if feasible, to match on certain initial race relations characteristics as well. Other kinds of characteristics such as region and ethnic composition will be controlled through the sampling design.

Wherever possible the two desegregated schools will differ on some characteristics which differentiates the desegregation policies of the two desegregated schools and which is felt to be important to the success of desegregation. Examples might be compulsory plans versus voluntary plans; percentage of minority students in the school; desegregation of faculty; classroom desegregation; the existence of a bilingual/bicultural program in segregated and desegregated schools serving Mexican-American and Puerto Rican children; etc.

There are some difficulties with certain plan variables. If one were to choose a compulsory desegregated and a voluntary desegregated school as a pair, then the

two schools would most likely be from different school districts, and we recommend that the SPS be limited to quartets, each within a single school district. The large-scale longitudinal study includes a number of quartets matched across district lines, but it is preferred that these quartets not be used in the paired study. Even if this did not make the matching on achievement and socioeconomic levels more difficult, it would mean that differences between the two schools would be confounded with school district or community effects. It might be more feasible, then, to use the two desegregated schools to contrast crucial policies that can differ within districts from school to school. The sampling design should then guarantee that both compulsory and voluntary school policies are represented in the sample.

The matching process would take place by examining all school districts within a particular sampling stratum (e.g., northern, urban, etc.; see Sec. 11) according to average socioeconomic and academic achievement level of the white and minority students. Quartets and quintets should be composed of schools whose average minority and white student characteristics resemble those for the stratum as a whole. This will prevent a tendency for regression toward the mean.[2]

The Panels

The first year of the longitudinal survey will include grades 3, 6, and 9. The panels used for the SPS must therefore be drawn from these baseline grade levels. In addition, the SPS will include a fourth panel for grades 1-3 which will be started in the second year of the main survey. We feel that students in kindergarden are too young for the kind of assessments required for cognitive and race relations changes. Those elementary schools selected for the grades 3 and 6 panels can be used for starting the grade 1 panel in the spring of Year 2. The panels are shown in Fig. 8.1.

These grade levels appear to be excellent choices for a comprehensive assessment of relatively long-term changes due to desegregation. The panels cover the main times of basic transition for each school level—such as the transition from elementary to junior high (6-9) and the transition from junior high to high school (9-12). The latter group will be especially useful for studying the college decision-making process. The 1-3 panel is equally important for showing changes from what is essentially the fundamental baseline—the entrance to school.

There are some other bonuses as well. The three-year design means that in 1978 there will be a duplication of grades 3, 6, and 9; this can be used to assess any overall changes in the school system itself (see the next subsection on cohort analysis). More important, the use of grades 6 and 9 means that there should be a number of schools whose panel of students will be changing from a segregated elementary or junior high school to a desegregated junior or senior high school. These are important target groups for the quasi-experimental design, since for rigorous cause-and-effect inferences it is necessary to have pre-desegregation baseline measures to compare with post-desegregation changes.

Cohort Analysis

The most serious methodological problem with longitudinal panel studies is the attrition rate. Over a three-year period it is quite possible that up to 50 percent of

[2] Regression toward the mean can seriously impair a matching study when individuals are the sampling units. School averages, however, are far more stable aggregate statistics and hence should be far less susceptible to regression effects when they are used as the basis for matching.

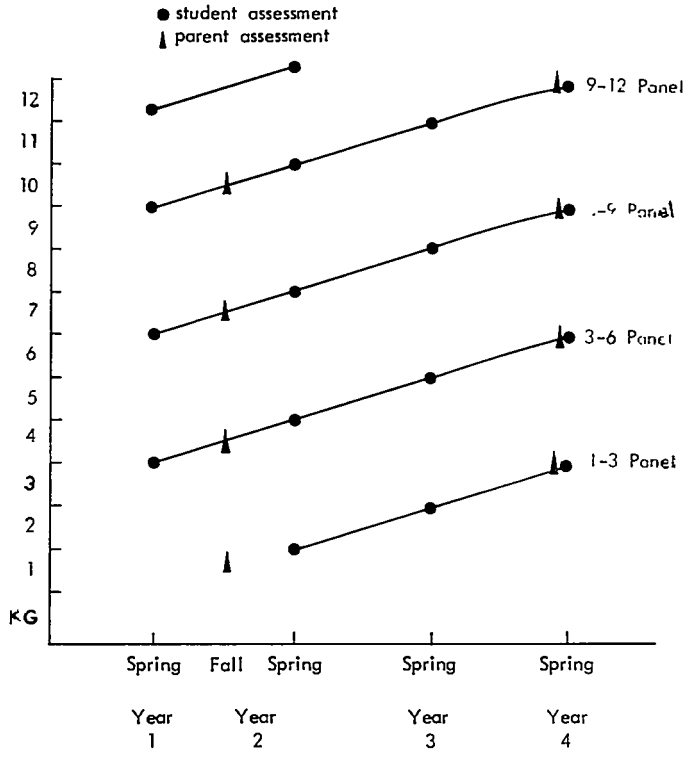


Fig. 8.1—Panels for the student panel study

the original panels will be lost through moving or changing schools, refusals to participate, and any number of clerical and administrative difficulties. For this reason it is important to assess either the entire grade level or a representative sample of students each year to replace students in that grade level who have dropped out of the panel. This should allow comparisons between the attrited panel sample and a cross-section of students for that grade to see in which way, if any, the attrition rate has affected its representativeness.

Parent Sample

Data from parents will provide basic family background information to be used primarily as intervening variables in the desegregation models phase of the investigation.

The basic plan is to collect information from all parents whose children are selected for one of the four panels shown in Fig. 8-1. Since panels cannot be selected until the first wave of main longitudinal survey is completed in the spring of Year 1, and since parent data should be obtained as close to the start of the student panels as possible, the first parent assessment will take place during the early fall (September or October) of Year 2.

A second parent assessment would take place at the end of the panel study in the spring of Year 4. This would enable us to use changes in parent attitudes both as intervening variables and dependent variables. Particularly in the race relations realm, the attitudes of parents are likely to be just as important as student attitudes for creating and maintaining a successful desegregation program.

Record-keeping and Confidentiality

A longitudinal study requires linking data from repeated assessments for individual students; this in turn requires that a student's name be connected with his/her questionnaire and test results. A system must be designed whereby the data can be linked to a name through some type of name-number file system, but so that neither the school nor the data base connects a particular name with a particular set of responses.

A system such as that used in the national longitudinal study can be modified to meet these goals. This requires a separate form for each student with his name, address, and names (and addresses, if necessary) of parents or guardians. This parental information is especially important for the parent interviews, but it can also be used for follow-up contacts to be made after a student leaves school (e.g., when he moves out of the school attendance zone). This form is also assigned a unique number for that student, and a master file of names and student numbers will be created. Next, all instruments used by the student (or any other respondent) are coded with some unique form number (not the student number) with a tear-off page containing his/her name, address, and the form number; the form number also remains on the instrument when the tear-off is removed. The tear-off sheets can be collected separately and merged by computer with the master name file in order to relate a student number to that particular form number; the student number then replaces the form number when the instruments in question are processed. In this way there is only a single, separate file which relates a name to a respondent number.

While these procedures may seem quite elaborate, it is important to stress the crucial need for confidentiality without sacrificing the ability to relate individual student data gathered at different times.

SAMPLING DESIGN

There are several levels of sampling that must be considered for the SPS. From the original set of school districts in the longitudinal study we must select matched sets of schools that are representative of desegregated schools in general. Second, students within schools will have to be sampled, and the sampling plans are likely to differ between elementary and secondary schools. Finally, some consideration must be given to sampling of parents.

School Sampling

The basic goal of school sampling is to obtain a small, representative sample of schools that are matched closely enough on most important student characteristics to qualify as a quasi-experimental design. In addition, the schools should illustrate the most important variations in desegregation plans because of their potentially central role in determining the success of desegregation. These conditions are easily met, since the SPS sample uses only 15 percent of the schools in the large-scale study. In order to accomplish both of these goals, therefore, it is important to carefully define the sampling strata.

The first stratum is *region.* The historical differences in the evolution of segregated and desegregated schools in the South demand a distinction between the North and the South. It is quite possible that models of successful desegregation will differ from the North to the South.

A second stratum is *type of school program.* For the Mexican-American and the Puerto Rican communities it appears that desegregation is generally seen as valuable only in conjunction with other types of educational intervention such as bilingual/bicultural education programs.[3] By comparing Spanish-cultured students in segregated and desegregated bilingual/bicultural programs to each other and to segregated Anglo and Spanish-cultured comparison groups it would be possible to assess the virtues of these various approaches vis-à-vis equality of educational opportunity.

Another important distinction is between large and small *sizes of cities or towns.* Much of the race relations turmoil of the past ten years has occurred in large northern cities; these areas may well have different outcomes than small cities or towns.

A fourth stratifying condition should be degree and type of *ethnic composition* of the school district (as opposed to the school itself). School districts with a large proportion of minority students tend to act differently than districts with a small proportion of minority students (e.g., compare the differences between Riverside and Pasadena, California). Three levels need to be distinguished: less than 30 percent minority; 30-50 percent minority; and over 50 percent minority (i.e., the minority group or groups have a majority of students in the public schools). The 30-50 percent group needs to be separated from the other two, since this is a composition level frequently associated with rapid turnover and with white anxiety about a transition to a minority-majority school system. In addition to level of minority enrollment, distinction must be made between different types of ethnic groups. The SPS will include both black and Spanish-American groups.

The final stratifying variable is *desegregation plan.* The community study will provide considerable data on plan differences; we anticipate that the following four different plans are most likely to affect the outcomes of desegregation:

Plan 1: Court-ordered. This means that a court has either decreed a desegregation plan or a school board has acted in anticipation of a court action. By definition most southern school districts fall into this plan.

Plan 2: Community-mandated. This category covers plans that are voluntarily initiated by a community through its school board without action by a court (or formally threatened action). The plans here are those which are mandatory for the entire district.

Plan 3: Voluntary. Voluntary plans are also initiated by a community, but they differ from Plan 2 in that the minority students are given a choice to attend a majority-white school or not, so that those minority students in a desegregated school are there by their own choice (or that of their parents).

Plan 4: Natural. Natural desegregated schools are those which, because of size and residential housing patterns, have been desegregated historically. These are important schools since they may tell us what to expect when communities desegregate residentially and not just educationally.

Using these strata, a sample of 20 matched quartets or quintets will be drawn for both elementary and secondary schools as shown in Table 8.1. Assuming quartets this will yield a maximum sample of 80 elementary schools for the 1-3, the 3-6, and the start of the 6-9 panel; and 80 junior and senior high schools for the 9-12 panel.

[3] See, for example, Sec. 5, Vol. I, of this report.

Table 8.1

SAMPLING QUARTETS OF SCHOOLS FOR THE SPS

District Ethnic Composition Percent Minority	North		South	
	Large City	Small City*	Large City	Small City
Over 50 percent:				
Plan 1 and/or 2	Q_B	X	Q_B	Q_B
Plan 3 and/or 4	Q_B, $2Q_P$**	X	X	X
30 to 50 percent:				
Plan 1 and/or 2	Q_B, Q_M**	X	Q_B	Q_B
Plan 3 and/or 4	Q_B, Q_M**	X	X	X
Under 30 percent:				
Plan 1 and/or 2	Q_B, Q_P**	Q_B	Q_B	Q_B, Q_M**
Plan 3 and/or 4	Q_B	Q_B	X	X

KEY: Q = quartet of elementary and secondary schools.
Q_B = black students are minority group.
Q_P = Puerto Rican students are minority group.
Q_M = Mexican-American students are minority group.
X = no schools sampled.

MAXIMUM SAMPLE SIZES: N = 80 elementary schools (for 1-3 and 3-6 panels),
N = 80 junior or senior high schools (for 6-9 and 9-12 panels).

NOTES: *Small city is less than 20,000 residents.
**The two quartets may be replaced with quintets.

It must be stressed that these numbers—and the entries in Table 8.1—are necessarily tentative, since the empirical distribution of schools within these categories will not be known until after the main longitudinal survey. For example, we have collapsed Plans 1 with 2 and 3 with 4 since we are not sure that both types can be obtained within the same strata (e.g., most court-ordered plans in the North have taken place in districts with a large proportion of minority students). Also, the placement of X's in the table may change depending on information from the longitudinal survey. Finally, without further information we cannot decide from which stratum to select quintets.

We anticipate that very few schools will change in ethnic composition markedly once the study begins. When they do change, it will be the result of resegregation, or white flight, from desegregated schools, or the implementation of a desegregation plan for a control school. It will be difficult to analyze those cases, and if any such marked changes occur, it may be necessary to exclude the school from some parts of the analysis. The study of changing schools is better handled by the longitudinal study of newly desegregated schools (Sec. 9).

The fact that a quartet consists of two desegregated schools means that differing types of desegregated schools can be matched within each quartet. Depending on the availability of schools and their similarity on academic and socioeconomic factors, the following comparisons will have priority: (1) Plan 1 with 2 and 3 with 4; (2) bilingual/bicultural desegregated and segregated schools with unilingual schools; (3) percentage of minority students in the schools (e.g., 15 percent versus 30 percent); and (4) extent of classroom desegregation. Obviously, more specificity is impossible until the first-year survey is initiated.

It would be ideal from many points of view to have each quartet or quintet drawn within the same school district. But this is ruled out for some strata by definition; many court-ordered plans will not yield any segregated control schools. Also, the constraints imposed by matching may rule out all segregated schools within certain districts. In any event, whenever quartets must be made up of schools from more than one school district, every attempt should be made to choose districts geographically near and with similar community and school policy characteristics except for those policies being contrasted.

The sampling will be determined for the elementary school; those junior and senior high schools that the elementary schools feed into are also selected.

Sampling Students

The large-scale longitudinal study student sampling is designed to meet the needs of the SPS. Let us briefly review the scheme.

The measurement of interracial contact and attitudes for the earlier elementary grades will require administration of a sociometric instrument of some kind with which a student can select those members in the class that he likes, dislikes, works with, and so forth. In addition, the measurement of teacher attitudes and expectancy will require that the teacher rank each student on various characteristics such as ability, motivation, etc. Both of these types of measurement require that all students in a classroom be included in the assessment. Further, since classroom assignments change from year to year, it is not possible to sample classrooms, since in subsequent years students from a given class will be dispersed throughout other classrooms.

Given these conditions of measurement it will be necessary to assess the entire grade level for each year of the 1-3 and 3-6 panels. While this will result in fairly large panels—perhaps up to 10,000 students in 80 schools—it must be remembered that attrition rates may be high in some areas. Moreover, this sample size becomes very small when basic partitions are made for ethnicity, region, social class, and basic plan variables. Within each of the basic strata shown in Table 8.1, for example, there would be only 400 students, and these would be further divided by ethnicity.

Since the classroom is not as meaningful a unit in the junior and senior high levels, and since the number of students in an entire grade is likely to be much larger at these levels, the 6-9 and 9-12 panels will be based on random samples of the grade level. For the 6-9 panel it makes sense to sample all classrooms in grade 6 so that classroom measures can be made at least for the first assessment; in subsequent years these students would then be tracked in whatever junior high school they attend. This will result in a panel of approximately 4000 students in 80 schools.

For the 9-12 panel a random sample of 50 students will be selected and followed through whatever high school they attend. In order to have comparability between the two ninth-grade assessments (one in the first year and one in the fourth), it will be necessary to sample 9th grade students in the 9-12 panel from the school attendance zone (if there is one) of the elementary school used to start the 6-9 panel. The 9-12 panel will also have approximately 4000 students in 80 schools.

In the final year of both the 6-9 and 9-12 panels it will be useful to draw a random sample of ninth- and twelfth-graders (from the elementary attendance zone) to replace those who move out of the school district or who are otherwise ineligible during the three years of panel assessments (estimated at about 30-40 percent). In this way we will have a random sample of the two cohorts corresponding to the panels, and therefore we can draw conclusions about the degree of bias in the three-year panel as well as about changes in the school system as a whole.

Those students who change schools during the three-year panel or who move out of the district will not be followed. Not only would it be expensive to do so, it is not clear how to use the data if they were available. Since we are assessing the effects of particular school desegregation policies, teacher attitudes, and so forth, the only group about whom clear inferences can be drawn are those who remain in a particular school exposed to those policies and teachers. Students that move to another school—even in the same district—will be exposed to other policies and teachers and thus any effects of the original school are confounded with effects of the new school.

Sampling Parents

While it would be useful to have parent data for all students in the panels, it would also be very expensive since there will be about 24,000 students altogether. Therefore, we anticipate that parents will be sampled in some way. One way might be to select some fraction of parents in the school panel, say 50 percent, but sample them in each school. Another way might be to sample 100 percent of parents but only for selected panels (e.g., 3-6 and 9-12), or perhaps for selected schools.

INSTRUMENTS AND DATA COLLECTION

The basic data collection process will consist of questionnaires, tests, and interviews administered to the students, teachers, and parents in the sample. Since the data collection instruments and procedures will differ for each panel or target group, we will describe each one separately. It should be noted that most of the data instruments will be in common to the main longitudinal survey, so a detailed description does not need to be given here. Further, the variables listed are tentative and not intended to be exhaustive; rather, they reflect those that should have high priority in a long-term study. The main study may measure a number of characteristics that prove unpromising or infeasible for a long-term study. A final selection is impossible, therefore, until the first wave assessment.

Panels 1-3 and 3-6

These groups are too young for any but the most straightforward of tests. Since aspirations, self-esteem, and race relations are difficult to measure at this age level, only the simplest instrument will be used. The first-grade instrument will contain only two areas:

(1) *Race relations*
Ethnic faces test (Gerard and Miller 1971)
Sociometric test

(2) *Academic*
Achievement test battery (reading; math)

The third-grade questionnaire is only slightly longer. The two additional items most relevant to the SPS are:

Self-esteem (third grade only)
Internal-external control (third grade only)

Panels 6-9 and 9-12

These two panels will receive the same battery of tests and questionnaires. While the sixth graders may be somewhat young for the full range of questionnaire items, it is important to obtain as much data as possible for comparison with later years. (It may be advisable to have instructions on the questionnaire cover telling them to skip any questions they do not understand.)

(1) *Race relations*
Ethnic faces test (6-9 only)
Sociometric test
Questionnaire
- Knowledge about other groups
- Race attitudes
- Attitudes toward desegregation, separatism
- Contact, positive and negative
- Anxiety

(2) *Academics*
Achievement test battery (reading in both English and Spanish where appropriate; math)
Questionnaire
- Job and college plans and aspirations
- Activities
- Attitudes toward school and teacher
- Study habits

(3) *Psychological and Other*
Internal-external control test
Questionnaire
- Social and demographic background
- Self-concept (academic, social)
- Ratings of happiness, optimism, trust
- Times disciplined
- Times in fights or arguments
- Attitudes toward Spanish and English Learning for Mexican-American/Puerto Rican, and Anglo controls.

Teachers

All teachers in an elementary school who teach in the grade level being assessed each year will be interviewed during the field visit to that school. For most elementary schools this should mean at most three or four teacher interviews. In secondary schools it may be necessary to draw a sample of 10 teachers from the pool who have contact with the sample of 50 students. The teacher self-administered form contains

basic information believed potentially predictive of student outcomes. We envisage three sections:

(1) *Questionnaire*
- Social and demographic background
- Professional training and experience, including proficiency in Spanish and English for teachers of Mexican-American and Puerto Rican children
- Interest in and satisfaction with teaching
- Attitudes toward desegregation
- Attitudes toward students learning Spanish and English
- Groupings beliefs and practices
- Traditionalist vs. modernist orientation

(2) *Adult verbal test* (similar to that used in *Equality of Educational Opportunity,* or a shorter variant)

(3) *Student ratings* (for each panel student taught)
- Academic ability
- Popularity
- Friendliness
- Discipline problems
- Language proficiency in English and Spanish (where appropriate)

In addition, the fieldworkers will fill out *interviewer ratings* on each teacher they observe:

- Warmth and empathy
- Degree of discipline and authority
- Time spent in instruction
- Support of desegregation

Multiple measures will be obtained in each school to permit inter-observer reliability tests.

Parents

Given the size of the parent sample, it would be cost-effective to conduct the parental assessment using mail questionnaires with telephone follow ups. Numerous new studies have obtained 75 to 80 percent return rates for mail questionnaires. The mailing could be done centrally, with the field team conducting telephone follow-ups where necessary. As a reliability check (and to meet the needs of the Community Study) 20 percent of the forms will be completed by personal interview. To shorten the instruments, alternate forms will be used, but all forms contain the following common core:

(1) Socioeconomic and demographic
(2) Aspirations for child
(3) Family size
(4) Race relations
- Race attitudes
- Knowledge

Support for desegregation, separatism
Contact with other races

(5) School-related
Involvement
Attitudes toward school policies
Attitudes toward teachers
For Puerto Rican and Mexican-American parents:
Attitudes toward their children learning
Spanish and English

School Data

The principal interviews done in the first and second year in the main longitudinal survey should be repeated in the last year to help measure changes in school structure and resources.

Summary of Sample Sizes

The design proposed in this section will result in the estimated sample sizes shown in Table 8.2 for students and teachers during *each year* of the study; the parents are interviewed only once.

Table 8.2

ESTIMATED SAMPLE SIZES

Panel	Students	Teachers	Parents
1-3	8,000[a]	300[a]	4,000
3-6	8,000	300	4,000
6-9	4,000	300	2,000
9-12	4,000	300	2,000
Total	24,000	1,200	12,000

[a]Excluded in the spring 1965 assessment.

DATA ANALYSIS

As a quasi-experimental design, the SPS requires data analysis techniques suitable to both change over time and control group comparisons. In addition, various other techniques are necessary for descriptive purposes and for constructing appropriate and reliable scales.

We want to stress that while the techniques described here reflect the current judgment of our design team, they do not comprise the only possible approach. Given the many different methodological strategies of different investigators, our hope is that the data collected in the SPS will be available to many researchers using many different techniques of analysis.

Reliability and Factor Analysis

Prior to assessment of the overall models, it will be necessary to construct certain indices and scales, particularly for the many attitudinal and personality constructs proposed in the previous subsection. Techniques of factor analysis with associated reliability coefficients are best suited for developing independent, coherent scales with satisfactory levels of internal consistency (Heise and Bohrnstedt 1971; Armor 1974).

In addition, the longitudinal nature of the design will enable an assessment of overtime and change-score reliability (Armor 1974). It is especially important to establish change-score reliability in the event that there are significant change effects for certain groups, so that factors relevant to change *within* the group can be identified.

Analysis of Variance and Covariance

The basic technique appropriate to a quasi-experimental design with more than two time periods is a two-way analysis of variance with repeated measures. One classifying factor is race or ethnicity; the other factor is either desegregation vs. segregation or, for testing models of successful desegregation, different desegregation plans. If a particular trend is posited for both races within desegregated schools but not segregated schools (or within one type of desegregated school but not another), then we would expect the time-by-desegregation interaction to be significant. A subsequent test for linear trend would also be appropriate.

Since it is not known at this time what kind of interaction might be present across different strata in the sampling design, it would probably be safest to conduct separate analyses of variance within the fourteen strata. If some stratifying variables have no apparent effect, then strata can be combined with a resulting gain in degrees of freedom.

In the event that the matching process does not result in reasonably comparable control groups, then a repeated measures analysis of covariance can be considered. It should be stressed, however, that most specialists in experimental design are of the opinion that analysis of covariance or any other regression technique does not fully correct for the original non-comparability of the treatment and control groups.

Levels of Analysis

The preceding discussion presumes that the student is the unit of analysis. Whenever appropriate, other levels of analysis will be considered. If classrooms are sampled at the elementary level, then the classroom aggregates might be computed and the unit of analysis could be the classroom. This would be a particularly appropriate approach for analyzing teacher effects. A similar case might be made for a school (or grade level) analysis, although with only 80 elementary or secondary schools there are not very many degrees of freedom.

Post-High School Effects

We believe that some of the most important effects of schooling will not appear until after high school graduation. It is important to see which schools encourage minority students to go to college or get good jobs, for example. We believe that this topic can be pursued most economically by examining data from three already existing studies, as described in Sec. 13. If it then seems wise to gather further data,

the initial eleventh-to-twelfth-grade large-scale longitudinal study could be used. A subsample of these students would be reinterviewed at the end of the SPS, when they would be two years past completion of high school.

Conclusion

The student panel survey is the only study proposed to follow students for more than one year. It is also the only study to link students as individuals to their parents. That means that this study is more effective than the others in four areas:

1. It is the most effective for studying the effects of school racial composition on students.
2. It is the most effective for studying the causal relationships between cognitive and noncognitive student characteristics.
3. It is the most effective for studying cumulative school and teacher effects.
4. It is the most effective for studying the effects of family background on students.

At the same time, the study has weaknesses. Its small sample size and the restrictions that matched schools come from the same district means that this study will not be able to consider all the interesting differences between schools and school districts. The requirement that schools have stable school populations means that we cannot use the student panel to study the initial effects of desegregation. The two types of studies proposed in Sec. 9 are an effort to address these issues.

9. LONGITUDINAL STUDIES OF THE DYNAMICS OF DESEGREGATED SCHOOLS

This section outlines a major set of substudies to grow out of the one-year large-scale longitudinal study. The basic design is to select a sample of schools from the larger study and carry out repeated one-year longitudinal studies of students in particular grades, enabling us to compare and contrast the growth of one group of students through the fourth, seventh, or tenth grade to the growth of the following years' students through the same grade(s). This serves two functions: it provides a replication of results (in, for example, measuring the effectiveness of a particular teaching style, or the effectiveness of a particular school innovation), and it provides an opportunity to observe the changes in schools as they become more or less effective over time. Much of this analysis will be done at the classroom level. This set of substudies will serve three main functions:

1. It will describe the dynamics of the school over the first few years of desegregation.
2. It will analyze the effects of desegregation and school innovations on teaching behavior, and study the impact of alternative teaching styles on student cognitive and noncognitive growth.
3. It will evaluate the effectiveness of innovations or intervention strategies designed to aid desegregated schools.

We recommend that three separate but related samples of schools be studied: a sample of newly desegregated schools; a sample of elementary schools containing promising innovative practices or interventions; and a sample of high schools with similar promising school characteristics. In Sec. 13 we present an optional approach to this set of studies based on the scientifically stronger but administratively more radical concept of using a randomized allocation of particular innovations and intervention strategies in order to permit evaluation with a genuine experimental design. If that strategy is not preferred, we propose that three samples of schools be used to generate three separate substudies as described below.

1. A STUDY OF NEWLY DESEGREGATED SCHOOLS

The first substudy focuses on a sample of up to 120 newly desegregated schools (equally divided among elementary, junior high, and high schools) to be observed over a period of years to examine how they change in response to desegregation and in response to efforts to improve the quality of education and the quality of race relations. The number of schools is limited by the number of newly desegregated schools in the main sample. Since the schools are subsamples from the sample used in the main longitudinal school survey, data exist for pretest and posttest scores for the fourth, seventh, tenth, and twelfth grades. In the subsample of schools used here, pretest and posttest surveys of the three lower grades (fourth, seventh, tenth) will be repeated for three additional years so that we can observe changes both in school inputs and school outputs. (We might call this a "second-order" longitudinal design, since each year is a data point which is itself the result of a longitudinal study.)

The main function of this study is a simple one: to describe the dynamics of the

adaptation of a school to the fact of desegregation. Desegregation is an exogenous "shock" to a school's traditional social and educational behavior in two ways. First is the fact of desegregation itself, which brings problems of race relations and community controversy which the school must learn to deal with. Second, desegregation brings a new set of students, who because they come from a different environment and occupy different roles in the community social system, bring different needs and demands to the school. We do not expect the school to adapt instantaneously to the situation. Rather we expect a period of "growing pains," sometimes minor, sometimes critical. We expect the school to make fundamental decisions during the first year of desegregation which may set the tone of race relations in the school for future years.

We have little previous research in this area. We have, for example, no theory of the stages through which a school must go in adapting to desegregation. Nor do we have any idea which of the many decisions made in a newly desegregated school are critical in making desegregation effective.

We recommend that this substudy be limited to desegregation of black and white students only. The number of newly desegregated schools with large numbers of Mexican-American and Puerto Rican students will necessarily be small, making any analysis of the way in which these schools react to desegregation difficult. The Mexican-American and Puerto Rican student panel studies (see Sec. 8) will provide an opportunity to deal with a number of the issues involved here as they apply to children of Spanish-speaking parents.

The study should pay special attention to the impact on students of administrative strategies, staff behavior, and community preparation. We can expect to find that some principals are highly capable of preparing the staff for desegregation and making the series of difficult decisions which are needed if the school is to transform itself from a merely desegregated school into an integrated one. The first few years of desegregation should show a change in the attitudes and behaviors of teachers as they react to the presence of a new group of students. In part, their reaction will be conditioned by their initial expectations and their racial and educational attitudes; but their responses will also be influenced by their in-service education and the nature of the administrative leadership of the school. Both the administrator and the teacher will be strongly influenced by the community, which provides a set of pressures and expectations that will cause them to react in different ways. We expect that an analysis focused on these three dimensions of the school will produce policy-relevant findings.

A longitudinal study over three years, with spring-to-spring data collected on students of the same age in successive years, is well suited to the task. It permits us to measure the change in a school itself, and to observe the extent to which that change results in increases in majority or minority cognitive or noncognitive growth.

A three-year study also serves two other functions. First, it permits us to replicate our findings in successive years. If a school with a particular structural arrangement of teaching, a particular kind of remedial program, or a particular use of equipment shows unusually high gains in one year, it is important that we know whether the same gains reappear in following years. Secondly, the three-year design permits us to successively refine our hypothesis and measuring instruments. We will begin with the findings from the large-scale longitudinal study, translating them into more sharply focused hypotheses, testing these, modifying the hypotheses further, and so on. Such a strategy provides a much greater chance of developing sound conclusions. The data collection is summarized in Table 9.1.

We propose that a mixture of survey and case study methods be used. The survey

Table 9.1

DATA COLLECTION PLAN--LONGITUDINAL STUDY OF NEWLY DESEGREGATED SCHOOLS

Year		
1	Pretest first cohort of students. Survey teachers, principals.	Large-scale Longitudinal Study
2	Posttest first cohort of students. Survey teachers, principals. Select subsample, pretest second cohort, in subsample.	Large-scale Longitudinal Study
3	Revise survey instrument. Begin school observations. Posttest second student cohort. Pretest third student cohort. Survey teachers, principals. Revise survey and observation instruments.	Subsample: newly desegregated schools
4	Posttest third student cohort. Pretest fourth student cohort. Survey teachers, principals.	Subsample: newly desegregated schools
5	Revise survey and observation instruments. Posttest fourth student cohort. Survey teachers, students. Complete school observation.	Subsample: newly desegregated schools
6	Final report synthesizing case study and survey results.	Subsample: newly desegregated schools

will provide reliability and accuracy of conclusions; the case study, richness of ideas and data. The typical problem of the case method is that data are gathered for only a single site, and the evaluative conclusions (Is this school more or less successful than others?) are often little more than speculation. At the same time, only the case study method has provision for the unanticipated discovery, the discovery of a radically new hypothesis, or the chance to observe the subtleties of organizational or personal behavior.

For these reasons we propose that survey methods and "comparative case study" methods be combined, wherein an observer will simultaneously study several randomly selected schools, testing his hypotheses against not only his or her own observations, but also against the survey data provided from student, teacher, and principal questionnaires and student achievement tests.

School Sampling

We propose that 15 to 25 matched pairs of elementary schools (i.e., the desegregated schools from the matched quartets described earlier) and the junior high schools and high schools which they feed be selected together. We recommend this because segregation is a districtwide process, and we are interested in analyzing the way in which particular district plans affect lower, middle, and upper grades. Ideally, one or both schools in each plan should have begun desegregation during the first

or second year of this study or in the two preceding years. (This means that at the end of the study we will have data covering each of the first seven years of desegregation.) There may not be 30 to 50 schools in the large-scale sample which desegregated in this four year period. Much depends upon the year-to-year fluctuations in the national pace of school desegreation. If there is relatively little activity, the sample may be as small as 12 pairs—a total of 72 schools in all. Where possible, districts should also be selected using the matching developed in the district-level survey. This will permit comparisons of schools within districts, using the sets of schools, and also comparisons across districts using the matched sets of districts.

While this sampling procedure will leave little freedom to the researcher for additional selection criteria, we nevertheless propose that in selecting districts and schools some additional criteria be considered to the degree that it is possible to do so. In matching districts with each other, it is useful to maximize the variation among districts in the *general racial liberalism of the school administration and school board,* since we predict that the schools will be strongly affected by the district's ideology, and wish to be sure that we have enough differences between districts to analyze this effect.

In selecting pairs of schools within districts, the critical variables are the racial composition of the school, the proportion of students who attend the school nearest their home, and the racial attitudes of the staff. We hypothesize that part of the widespread objection to busing—especially from parents of bused students—is in fact an objection to students attending a school in a residential neighborhood that is foreign to them. Viewed in this way, we see that neither the distance from the student's home to a school nor bus riding is the controlling factor—rather it is the nature of the relationship between the student, his or her neighborhood, and the school. The sample should provide contrasts between schools in the percentage of students of one ethnic group not attending the school nearest their home.

As many pairs of schools as possible should provide contrasts in their racial composition; comparison between schools where the two groups are nearly equal and schools where one group predominates are especially interesting. We anticipate that a large number of relationships between racial composition and student outcome will be curvilinear; for example, we expect that both the level of positive racial interaction and the level of negative racial interaction will reach a peak in schools which have 50-50 ratios, falling off as either ethnic group becomes predominant.

The final variable that should be considered in the selection of pairs of schools is the attitudes of the staff toward desegregation. Since we anticipate that the effects of this variable will be very powerful, making sure that we have some pairs of schools that are matched on staff attitudes will enable us to better control this effect in order to see other, more subtle ones.

The sampling procedure is as follows. All schools that desegregated in the first (or second) year of the study, or in the two previous years, are listed, along with the other desegregated members of their set. The pairs are then stratified into North and South, and within each region, dichotomized according to the racial ideology of the superintendent and board; this dichotomy is from the community district survey, based both on their private expressions of opinion and the opinions expressed by other black and white community leaders about them. No more than 4 to 6 sets of matched elementary schools and the secondary schools they feed are selected from each of the four strata. Each pair is sampled with a probability proportional to the number of "yes" answers to the following questions:

- Do the two elementary schools, the two junior highs, and the two high schools all differ in racial composition by a moderate amount (e.g., 20 to 40 percent)?

- Do the schools in a set differ in the percentage of students of each race who are attending the school nearest their home?
- Are the schools in a set similar in the racial attitudes of their staff?
- Will the fourth, seventh, and tenth grades all be affected by the newness of the desegregation plan during the next three years? (For example, will the age at which tenth-graders first experienced desegregation be different for the next three cohorts?)

In selecting pairs of school districts, data from the screener survey, the district survey, and the first wave of the large-scale longitudinal study can be used to verify that student background characteristics in the comparison schools are indeed similar. However, the criteria for comparison should not include any of the outcome variables from the school study. If we match two schools with different economic backgrounds but similar achievement test scores, we will be unable to determine what factors caused the lower income school's scores to equal those of the higher income school.

Survey Data Collection and Instrumentation

The data collection for various grade levels at various times is indicated in Fig. 9.1. The basic concept is to repeat the large-scale, one-year longitudinal study on three successive cohorts of fourth, seventh, and tenth grade students. While it would be possible to carry forward a twelfth-grade series of studies, we recommend that this not be done; the tenth-grade panel is preferable to the twelfth since in many of our schools the students will be experiencing their first year in the new high school, and school effects will be strongest for them at that grade, making the analysis of school effects easier.

The sampling of students is identical to the large-scale longitudinal study: the entire third-grade class is surveyed, and in sixth and ninth grades a sample is drawn containing (if there are sufficient students) 35 students of each ethnic group living

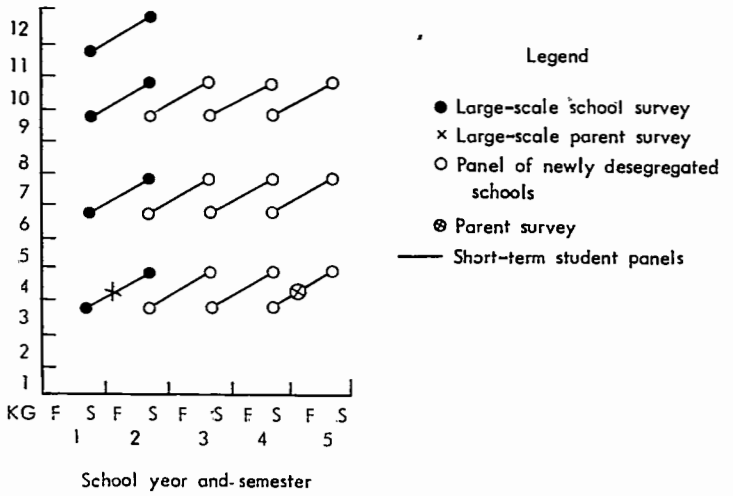


Fig. 9.1—Grade levels to be surveyed in panel of newly desegregated schools

in the elementary school attendance zone plus additional students randomly sampled from the remainder of the school. The secondary school surveys are to be carried out in English classes, and the questionnaire modified in order to take advantage of the fact that we have data on a group of students who are interacting together in English classes. English classes were selected for intensive study in part because this is a core section of the curriculum which most students will be taking, and also because the nature of the English curriculum provides opportunities to introduce racial issues into class discussion in a way that math and science classes do not. This means that the racial behavior of English teachers, and the racial interaction in English classes, will be disproportionately important in setting the racial tone of the entire school. (History or social studies classes could be used for the same reasons, but differences between districts in curriculum would make it difficult to draw comparable samples.)

In the fall of year two, fourth-grade parents are to be interviewed in the main school survey. We recommend a resurvey of parents in the fifth year to measure change in parent attitudes toward the school.

The instruments to be used in the school survey will remain very similar to those used in the earlier large-scale longitudinal survey. Additional items should be added to measure in more detail staff perceptions of race relations, the introduction of new programs, and the perceived change in the school as reported by all participants. Items dealing with parental involvement and relationships of the school to the "sending" neighborhoods should be added, and the section dealing with qualitative racial interaction should be strengthened since race relations is the main focus of this study to an even greater degree than it is in the others.

Because there are relatively few opportunities to drop questions from the original longitudinal school survey, the researcher is likely to retain almost the entire original instrument. The researcher should be provided funds to permit adding questions each year to pursue new hypotheses as they develop from the analysis.

Case Study Data Collection

The survey should be supplemented by considerable classroom and school observation, partly to validate the key measures of the study—such as levels of racial contact, racial tension, teacher interaction with parents and students, etc.—but more importantly, to provide us with material with which to explain the findings generated from the survey. Consider, for example, the possibility that we find some school variable associated with several outcome variables in a complex fashion, showing a positive association with some outcomes and negative association with others. It will then be possible to generate complicated hypotheses to explain these results, but it would be very unlikely that the survey instrument would have the particular questions necessary to test them. Of course, those additional questions could be added in the following year. But meanwhile, observational data on the school could provide opportunities to pursue some of these hypotheses. This means that we do not need to depend on observational data to provide "hard" measures of student attitudes or behavior or the other key variables in the study. Rather, what we need is the opportunity to develop new variables as the need arises. This means that the observational procedure should be very wide-ranging, covering as many different aspects of the school as possible and recorded so as to permit all aspects of those data to be recovered. We therefore recommend that the data not be reduced to computer-readable format, but be retained in the form of notebooks and tape recordings. The observer teams should use categorical responses and answers to open-ended questions to provide a brief (5 to 15 pages) summary of each case. These

summaries should be prepared independently by each member of the team to provide a reliability check. These then provide a reference which can be used during the analysis of the survey data.

Observational data can also be used to generate new hypotheses in advance of survey work. For example, if observers are sent on successive trips to schools which varied in the number of years of desegregation they had experienced, the observers' overall impressions of the way in which districts changed over time may be very valuable in developing questionnaire items for the survey.

The use of observational data in the longitudinal framework would also provide opportunities for observers to report the quality and nature of change that they see occurring.

We propose that during the first longitudinal study of nearly desegregated schools (in year three of the project) that eight person-days of observation time be devoted to each of twenty high schools, twenty junior high schools, and twenty elementary schools selected from this study sample. This would enable six teams of two persons each, covering one school each week in the field, to do all the observation over a period of less than four months, allowing for time in the office transcribing notes, etc. (We realize that many researchers would recommend more time, but we believe this is sufficient)

The observers should use semi-structured interview forms as a guide to their field work, with general agreement in advance on the number and kind of respondents to be interviewed and the types of situations they should observe. Considerable emphasis should be placed on observing students in out-of-classroom activities, such as playgrounds, cafeterias, schoolgrounds before and after school, gym classes, etc. The teams should be debriefed with tape recorders and should interact with other teams as frequently as possible in order to share ideas. The whole logic of the observation is developmental—the observations in the last schools will be considerably different from those in the first schools, as new variables are defined and new observation techniques developed. Because of this, the observers should participate on a continuing basis in the design decisions. During the first year, the goal of the project is to search for new hypotheses and to provide a written record which can be referred back to during the analysis of the survey data.

During the second and third years of the longitudinal study of newly desegregated schools (years four and five of the project), the purpose of observation will change slightly, to provide data with which to clarify hypotheses developed during the first year's analysis. This means that the interviewing may involve more schools. The schools to be observed will be selected on the basis of the analysis of the preceding year's survey data.

Finally, during the last year of the study, the methodology should change considerably. Since at this point exploratory hypotheses are of considerably less value, the observations should be changed to provide highly structured "hard" measures of certain variables which the survey has failed to measure adequately in the past and which are necessary to test certain hypotheses.

Analysis of Survey Data

In measuring the changing effectiveness of the schools we are primarily interested in seeing how the slopes of the "growth curves" for various student outcomes change from one year to the following years. Figure 9.2 presents an example of what these measurements might look like for a variable that increases monotonically from the beginning of elementary school through the sixth grade. In this case the post-test scores on the variable (here we use internal vs. external locus of control as

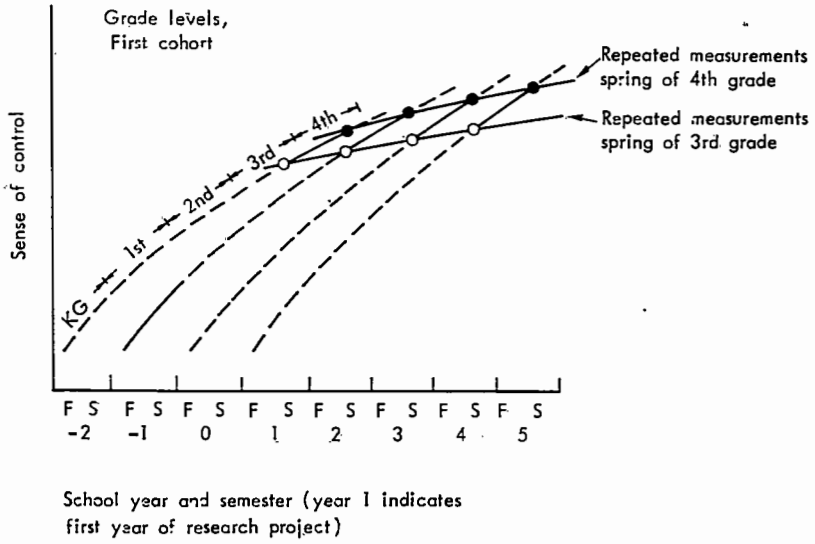


Fig. 9.2—Plot showing effects of increasing quality of education, reflected in steeper growth of student sense of control of environment

our example) are always higher than the pre-test scores; in addition, our example school shows steadily increasing growth for each successive cohort over the four years of the study. Thus the students who entered the fourth grade during the second year of the study begin with slightly higher control scores and the slope of their growth during the fourth grade is higher, so that the gain is slightly larger at the end of the fourth grade when they are compared to the preceding cohort.

When a variable is used which does not show a monotonic growth curve, the analysis is similar, although the results are difficult to interpret.

This analysis procedure will permit this study to assess three things: the degree to which desegregated schools change during the first few years of desegregation; the degree to which this change is reflected in changes in student outcomes; and finally, the way in which each cohort is affected by the increased duration of a newly-instituted program. Thus, if our schools have all desegregated within the past three years, the initial cohort will not have begun experiencing desegregation in kindergarten, while later cohorts will have, and this study design will permit us to assess these effects.

Use of an elaborately developed matched-pair design is not meant to imply that analysis of covariance is the only, or even the preferred, method of analyzing the data. Conventional regression methods, pooling the entire sample, can be used. The only serious problem is that the sample selection will cause the standard deviations of certain variables to vary considerably from those produced by a random sample, and therefore the use of standardized regression coefficients might be quite misleading. Similarly, techniques that use percentage of variance explained are subject to the same problem. However, an analysis using unstandardized regression coefficients would be quite appropriate.

In addition to conventional regression techniques, it is possible to use an analysis of variance design with districts blocked into matched pairs. This provides the

best possible measure of the statistical significance of a particular result, but is a generally poor procedure with which to search the data looking for interesting results. A compromise strategy between the two would be the construction of a data tape in which a pair of cases is the unit of analysis. This would permit an analysis in which the difference between the two schools on one variable could be correlated with a difference on a second variable. Such a procedure would take advantage of the match-pairs design and would be appropriate if the analysis of covariance indicated that blocking had a significant effect on the outcome variables.

Analysis of data should not be limited to a small number of hypotheses and during the first two years should be viewed as largely exploratory, with no effort made to produce definitive final statements until data from all years are available. The researcher should begin analysis during the third year of the project, using data from the large-scale longitudinal study to search for hypotheses. As each additional block of data becomes available, analysis of these exploratory hypotheses can be refined.

A number of causal models should be tested in an effort to locate as many intervening variables as possible between student output variables and staff input factors. As each model is developed, it should be tested against the case study material. In many cases, it will be useful to have the case studies reread to obtain computer-recordable judgmental scores on certain variables so that the case study data can be incorporated directly in the statistical analysis. (This should not be done until the analysis has proceeded far enough to determine the precise definition of the variable of interest. Coding of the case-study data should be done by judges who are ignorant of the outcome scores for each school and the hypotheses being tested.)

In general, the analysis topics should be guided by, but not limited to, the model developed in this report. The model should be supplemented by additional theoretical work dealing with the topic of how school systems react to exogenous shocks (in this case, in the form of desegregation) and how they respond to pressures from the community, their students, and parents. Detailed data on the process of decision-making within the school would be quite valuable in connection with the study.

2. A LONGITUDINAL STUDY OF STRATEGIES FOR SUCCESSFUL DESEGREGATION OF ELEMENTARY SCHOOLS

We propose that the large-scale longitudinal study data base and the existing research literature be used to select three programs that seem effective in elementary schools in improving either achievement, race relations, or emotional outcomes, and subsample the schools that have the most highly developed versions of those programs. These schools, along with a control group of desegregated schools, would constitute the subsample to be used in two repeated one-year (spring to spring) longitudinal studies of the fourth grade to determine whether the programs in question do in fact produce results with regularity, what particular aspects of the program are important, and what social process or theory of education explains their effect.

In selecting programs for intensive study, there are two alternative strategies. We may select programs a priori, because they are intrinsically interesting, or we may select programs empirically, on the basis of evidence of success. We propose that two programs be selected a priori and a third empirically. The two programs we believe deserve further study are (1) the intensive use of individualized instruction and (2) the intensive use of a multi-ethnic curriculum. Both individualization and

the multi-ethnic curriculum represent leading edges of a humanistic movement in the public school system. We would predict that they or some variant would continue to gain support many leaders in education. For these reasons, a detailed study is worthwhile.

The argument for individualization has not been highly developed with special reference to desegregated schools, but the way in which individualization poses a solution to the problem of heterogeneity of achievement levels in desegregated classrooms has been pointed out frequently. At the same time it has been argued that the lack of structure frequently accompanying curriculum individualization works to the detriment of students from poor families. Thus this topic is relevant, theoretically interesting, and the results of the study worth learning. The multi-ethnic curriculum has been the subject of much discussion but relatively little research of high quality. Again, given its growing popularity and the lack of consensus about either its theoretical basis or its effects, the topic is a worthwhile one for specialized study. In the course of undertaking detailed studies of these two topics, the research group should be asked to develop a third topic for extensive analysis, based on their conclusions drawn from the large-scale longitudinal study.

After this third topic has been chosen and the longitudinal study analyzed to develop a reconceptualization of individualization and multi-ethnic curricula, schools exhibiting highly developed forms of these three types of programs (whether or not they show benefits for students) should be sampled along with control schools, and studied with a spring-to-spring longitudinal analysis of fourth-grade classrooms for two repeated years. In the course of this analyis, the research team should pay special attention to three additional variables: the characteristics of teachers and the character of teacher-student interaction in the classroom; the role of heterogeneous vs. homogeneous ability grouping; and the behavior of the school's principal and other administrators. While the sampling scheme and data collection procedures are similar to those discussed in the first part of this section, it should be borne in mind that the basic logic is very different. Whereas in the preceding substudy particular school or district characteristics were used as guides to the sampling, in this study the selection of schools exhibiting certain types of programs is mandated. Secondly, the preceding study used three successive longitudinal studies in order to observe change in the school; in this case we propose two longitudinal studies of fourth grade students to verify that the program has consistent effects.

The Sampling Procedure

The first stage in the sampling procedure is the analysis of the large-scale longitudinal study to prepare a preliminary set of results early in the fall of the third year of the project. These results should be in the form of measures of effect (such as regression coefficients) for a large number of different school characteristics assessed against a number of school outcomes. The listing should be biased in favor of those programs most obviously manipulatable by outside policy-makers, but not restricted solely to such programs. The research team conducting this substudy will use these data to choose the third intervention strategy to be tested and will themselves pursue further analysis to clarify the type of individualization program and multicultural curriculum they consider worthy of further study. In the course of the analysis, the research team may propose dropping either individualization or multiethnic curricula in favor of a fourth topic.

After the topics have been selected and refined, those schools with the most highly developed programs of each type should be sampled. We anticipate that perhaps 10 percent of the total sample will have programs developed to the extent

that the research team will consider them adequate for evaluation. If three topics are used, this would yield a sample of at most 78 schools, since there are 260 desegregated elementary schools in the large-scale study. (In fact, there will be some overlap among the three programs.) A sample of control schools is next selected, consisting of a one-third sample of the matched desegregated schools from the quartets that furnished the schools whose programs are under study. This yields a final sample of at most 104 schools.

We anticipate one very serious problem in the sampling. We would assume that those schools with highly developed special programs will have unusual staffs, since it is likely to be the commitment of the principal and his staff which caused the school to be unusual. This means that what appears to be a program effect may instead be a staff effect, and problems of contamination will be difficult to solve. We may exaggerate the seriousness of this problem, and the issue can be resolved by examination of the large-scale study results; but if the problem is as serious as we think, there may be no satisfactory solution short of an experimental design such as we propose in Sec. 13. Short of that, it may be useful to oversample control schools without any of these special programs but with teacher characteristics similar to those in schools that have highly developed programs of the three types under study.

Data Collection

Unlike the longitudinal study of newly desegregated schools, analysis of the large-scale school study will require a one-year delay in the data collection from subsample schools. The sample can be chosen in the winter of the third academic year of the project.

Data from students, teachers, and principals are gathered in the spring of years three, four, and five, as shown in Fig. 9.3. In year three, only third-grade students are surveyed; in year four, these students (now fourth-graders) are resurveyed along with a new group of third-graders; and in year five, the second group (now fourth-

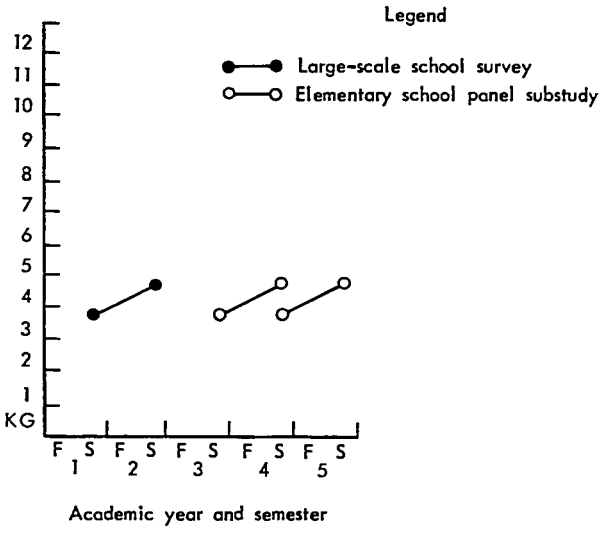


Fig. 9.3—Grade levels and survey dates in panel of elementary schools

graders) is resurveyed. In each year, third- and fourth-grade teachers are surveyed, providing both pre-test data on fourth-grade teacher behavior and attitudes, pre-test data on third-grade teacher behavior and attitudes, and three-year longitudinal data on behavior and attitude change among teachers. The spring pre-testing is an important element of this design, since it has been pointed out that teachers are influenced by students, just as students are affected by teachers, and it is therefore valuable to gather pre-test data from both groups before they come together.

In terms of timing, the critical period is the third academic year, since the final decision must be made about programs, schools selected, and data collection instruments prepared in time for a May administration of pre-test data.

Data to be collected will consist of all basic longitudinal school survey instruments, plus considerable additional survey instrumentation and observational data collection. Additional data collection should cover detailed descriptions of special programs, including descriptions of the projects and measurement of various intervening and dependent variables as needed to test alternate hypotheses about why these programs are effective or what conditions are necessary for their effectiveness.

In the case of individualization, special attention should be paid to the use of electronic media.

Additional data should be gathered on heterogeneous and homogeneous grouping of students. Particular attention should be paid to the opportunities for students to interact with students outside their own classroom, through the use of students moving outside their classroom during certain parts of the day, the use of extracurricular activities, and also procedures for reassigning students from one achievement group to another. Attention should also be paid to the use of achievement grouping in classrooms.

Additional data in administrative and teaching behavior should also be gathered. Here the special emphasis should be on measures of reported and perceived behavior—the frequency the teachers report for different kinds of activity, the perception of teacher activity by students, and similar data about various administrative behavior on the part of the principal, both self-reported and as perceived by teachers.

Classroom Observation

This study has as its focal point behavior within classrooms. While student reports of classroom climates and of the behavior of their teacher are valuable, problems of respondent bias are obviously difficult to deal with. Problems of teacher self-reports are, if anything, more serious. For these reasons we propose that classroom observation data be gathered in each school. These observations can be made by members of the survey team if they are given special training in using the interaction process analysis type methods proposed. In the last part of this section, we discuss classroom observation in detail.

Observations should also be made of the work of the principal, although without highly developed instruments. Here a simple scoring system to record the percent of time devoted to various types of activity may be most valuable, coupled with judgmental scores of the quality of interaction between the principal and his staff.

Insuring observer agreement on the procedures to be used and on the coding categories is a major problem, which should be dealt with by providing for intense interaction between observers after each field trip.

Judgmental ratings by the survey team can be subject to straightforward reliability checks; if reliability of data can be demonstrated, these may prove to be some of the most valuable material. (One procedure is an analysis of variance, to demon-

strate that the between-school variance is large relative to the between-observer variance, which is the error due to the lack of reliability).

The Analysis

The three basic tools are (1) analysis-of-covariance type measurement of the effectiveness of alternative programs compared to each other and to the control schools; (2) correlational analysis to verify that growth rates of students in different types of schools are indeed stable over the three cohorts of the large-scale study and this substudy; and (3) regression and path analysis to locate the intervening variables that link a program's characteristics to the outcome variables. As in all of the studies, the full list of student outcome variables should be analyzed. In addition, we expect that teacher attitudes and behavior may be strongly affected by the programs, so these are important both as intervening and dependent variables. In analyzing teachers, attention must be paid to the fact that the sampling procedure may have produced a sample of unusually gifted teachers.

Analysis of the impact of school characteristics should begin in year three, using all available data to begin testing hypotheses about school effects. Working papers should be widely circulated in order to provide maximum opportunity for modification of the hypotheses to be tested with later data collection. The analysis should use not only the data from this substudy, but also the data from the large-scale study, the student panel, and the study of newly desegregated schools in order to replicate findings from this study and develop new leads for explanatory hypotheses.

3. A LONGITUDINAL STUDY OF STRATEGIES FOR SUCCESSFULLY DESEGREGATING HIGH SCHOOLS

We propose that a study of high schools be conducted very much along the lines of the preceding study of elementary schools. The two designs are identical in terms of strategy of sampling, dates when schools will be surveyed, and method of data analysis. The only differences are in the area of programs to be studied, dependent variables (school outcomes) to be measured, and use of observational techniques.

While the studies have very similar designs, we expect that results of the analysis and the basic theoretical views used in explaining the data will be quite different, since adolescents and pre-adolescents differ considerably from elementary school children.

Programs to be Studied and Sampling Scheme

We propose that both junior high schools and high schools be studied, focusing on the seventh and tenth grades. Recall that in the elementary school design the sampling was done to maximize variance on three variables, two to be selected a priori and the third selected from the results of the large-scale longitudinal study of schools. We propose that the same sampling plan be used here except that the programs be changed to emphasize out-of-classroom activities for students and the use of human relations programs.[1]

In the case of extracurricular activities, we are not referring only to the customary cluster of clubs and sororities, but also to programs in which groups of students

[1] Both of these topics are discussed in Sec. 4, Vol. I, of this report so that it is not necessary to elaborate on the theoretical issues here. There are, however, serious problems of operationalizing the concepts.

perform and develop skills in art, technical areas, community service, athletics, music, and drama, and in particular to programs where student groups perform so as to bring credit to their school. Thus we are concerned with structured task-oriented activities rather than a network of social groups, and we are especially concerned with ways in which out-of-classroom activities contribute to the creation of a sense of school pride. No data exist to tell us whether these factors tend to be positively correlated in the universe; it may be possible to sample schools with varying types of athletic programs and schools with differences in musical and dramatic and other types of activities separately.

The selection of schools with extensive human relations programs is less a problem of conceptualization and more a problem of measurement. Many schools will claim to have highly developed human relations activities. Actual behavior measures of participation in these programs and some description of the content of the program will be necessary in order to select schools of interest to us. These data will all be available as a result of the first-year longitudinal school study.

The proposed third program to be oversampled, to be located on the basis of measures of effectiveness generated from the first-year study, can be selected in much the same way as was proposed for the elementary school design.

School Outcomes for Secondary Schools

Schools serving pre-adolescents and adolescents must deal with aggression problems that are generated by students of this age group. School race relations become a much more serious problem in junior high school than in elementary school, partly for these reasons. We recommend that the high school and junior high school studies focus considerably on problems of aggression and violence, both of majority groups against minority students and of minority students against the majority group. The greater mobility of older students also means that choice of friendships will be less bound by such school-control factors as classroom placement or seating arrangements in classrooms. Students will be free to choose friends from a large portion of the school. For these reasons, sociometric data will be extremely valuable.

Sampling of the Classrooms

We again propose selecting students in English classes, and aggregating data to the English classroom level. Questionnaires should be modified so as to focus some data collection on interaction with the English teacher and interaction in the English classroom, as well as to stress those outcomes such as achievement in grammar or literature that are most influenced by the English class.

However, we cannot emphasize the interaction of a single teacher with his or her students to the same degree that we did in elementary school. The important behaviors that are the basis of school social climate do not occur in the classroom; they can be observed more easily on the baseball field, in the lunch room, and at the school dance. We therefore propose that the systematic classroom observation techniques used in the elementary school study be replaced in part or entirely by more impressionistic observations of student interaction in a variety of situations. This is not to say that these data should not be machine-readable. We think that systematic data collection forms, and systematic procedures for permitting observers to score schools on overall characteristics of student body interaction and teacher-student interaction, will be quite amenable to conventional statistical analysis. As in the elementary school case, the design and analysis should be guided by, but not limited to, the model developed in this report. The literature on social relations in

high school is rich and valuable, and a few studies of racial interaction in high schools provide important sources and hypotheses.[2]

CLASSROOM OBSERVATION METHODS

Student and teacher reports of what goes on in the classroom are valuable, but these reports are subject to respondent bias; therefore, we propose that classroom observation data be gathered in each school (1) to study the behavior of teachers and pupils and (2) to determine whether teachers expect different performances from different students, and, if so, how these differential performance expectations are communicated by the teacher. Classroom observation would provide data on the resources available to the classroom and how these resources are used. First-hand observation would provide an overall impression of the climate of individual classrooms and of each school in general. For these reasons, we recommend that in the study of elementary schools with special programs, every fourth-grade class be observed for one day.

Numerous classroom observation schedules have been developed, each with particular advantages and disadvantages. The purpose for which classrooms are being observed should dictate the observation procedure. For example, one objective of classroom observation proposed here is to determine whether teachers communicate differential performance expectations to students. To achieve this objective, we need to look at what the teacher communicates, as well as at what the students communicate. Exchanges between student and teacher need to be observed.

We also recommend that a fairly low-inference model be used. Low-inference categories require the observer to note the occurrence or nonoccurrence of an event, whereas high-inference categories require the observer to interpret behavior. High-inference categories are subject to a great deal of observer bias, and if such measures are used we recommend they be used in conjunction with low-inference measures and be regarded as supplemental data.

In our review of classroom observation techniques, the Good and Brophy (1969) Teacher-Child Dyadic Interaction schedule appears to best meet the requirements of the objectives outlined above. The Good and Brophy system applies only to those classroom interactions in which the teacher is dealing with a single, individual child; situations in which the teacher is addressing the entire class as a group, such as lecturing, are omitted. This procedure is specifically applicable to studies that focus on intra-class individual differences.

Every interaction between teacher and individual child is coded. In addition, several aspects of the system involve preservation of the sequential nature of teacher-child interaction, so that cycles of initiation and reaction are not lost in the coding process. This feature is especially important for studying the communication of performance expectations, since it allows separation of effects due primarily to the teacher from effects due primarily to the child. This system also allows for conversion of raw codes from the individual children into percentage scores which neutralize the effects of differences in the absolute frequencies of various types of interactions they have with their teacher. Teachers' interactions with particular children or subgroups of children may then be compared directly with interactions in equivalent situations with other individuals or groups. In this way, *quality* of contact (what the teacher does when engaged in certain kinds of interactions with the child) and

[2] In particular, see Coleman (1961), Stinchecombe (1964), McDill et al. (1966), and Patchen et al. (1973).

quantity of contact (the sheer frequency of the different kinds of interactions) may be studied and evaluated separately. Finally, data for the entire class treated as a group may also be obtained by combining codes for individual members.

A more complete description of the type of classroom observation we propose follows; sample forms are in the Appendix.

SUGGESTED DESIGN FOR CLASSROOM OBSERVATION

Prior to scheduled observation, the following information should be collected from teachers whose classrooms are to be observed: a seating chart, showing the students by name; the ethnicity of the students; and the teachers' rankings of students according to achievement level. It would also be advisable to obtain a typical daily activities schedule from the teachers so that there would be a means for judging whether or not the activities observed were typical for that classroom.

Some of this information will be collected from the teacher survey, but other information, such as seating charts, will need to be gathered separately. The research staff could then preselect students for observation. Four to six students per classroom should be selected—two or three students from the upper third of the class in terms of achievement level and two or three students from the lower third of the class. Each pair or triplet contains students of different ethnicities so that in the analysis the effects due to race can be separated from the effects due to level of achievement. Schools that track students according to ability level should be identified, since the absolute differences between high-achieving students and low-achieving students will be less in schools that track. In case one or more of the students is absent the day the observing is to take place, student alternates, matched by ethnicity and achievement level, should also be selected. The seating chart and the names of the students to be observed, along with the names of the student alternates, would then be given to the observer. The observer should not know that students have been categorized by achievement level. The observation team would then be given training in using the observation method selected; from their experiences, Good and Brophy recommend a one- to two-week training and practice period to establish sufficient intercoder reliability.

The Good and Brophy System

Five different types of dyadic interaction situations are coded in the Good and Brophy system:

1. *Response opportunities,* in which the child publicly attempts to answer a question posed by the teacher.
2. *Recitation,* in which the child reads aloud, describes some experience or object, goes through arithmetic tables, or makes some other extended oral presentation.
3. *Procedural contacts,* in which the teacher-child interaction concerns permission, supplies, and equipment, or other procedural matters concerned with the child's individual needs or with classroom management.
4. *Work-related contacts,* in which the teacher-child interaction concerns seat work, homework, or other written work completed by the child.
5. *Behavioral contacts,* in which the teacher disciplines the child or makes individual comments concerning his classroom behavior.

These five broad categories of teacher-child interactions are kept distinct from one another in coding, and each type has its own place for coding on the coding

sheets (see Appendix for examples of modified coding sheets). In addition to this physical separation of the coding for the five types of dyadic contacts, coding distinctions are also made concerning the nature and sequence of the interaction observed. For every interaction, coders note whether the initiator was the teacher or the child, and also code information concerning the teacher's message or response to the child during the interaction. In addition, the coding of response opportunities and recitation turns also includes information concerning the type of question asked and the quality of the child's response, both of which are coded before coding the nature of the teacher's feedback. The latter coding also includes preservation of the sequential order of events, so that the chain of action and reaction sequences within these interactions is maintained.

Response opportunities are characterized by three key aspects:

- They are *public* interactions between the teacher and only a single child at a time, but are nevertheless meant for, and monitored by, the entire class or by the entire group operating at the moment (such as a reading group).
- They occur when the *teacher asks a question* demanding a verbal response from the child or when she asks the child to *publicly* respond to a question requiring a nonverbal response.
- Only a *single individual child* makes the response.

Response opportunities as used in this system are considered to be teacher afforded; they thus involve individual recognition of the child by the teacher. Each response opportunity which is coded requires coding of five separate bits of information: the identity of the child, the type of response opportunity, the level of question asked, the quality of the child's answer, and the nature of the teacher's feedback response.

Reading and recitation turns differ from other response opportunities in that the child is required to make an extended oral presentation, and the amount of teacher-child interaction expected is dependent on the performance of the child. The student who successfully completes his entire turn without error will ordinarily interact with his teacher only at the end, or when he or she makes some comment about his performance as a whole. The child who frequently makes mistakes along the way can expect the teacher to react to him each time he makes a mistake or gets stuck. Therefore, there is a separate reading and recitation coding sheet (see Appendix).

The last three dyadic teacher-child contacts in this system differ from response opportunities and reading and recitation turns in that the teacher is dealing privately with one child about matters idiosyncratic to him rather than publicly about material meant for the group or class as a whole. The latter distinction is the key one, since these teacher-child dyadic contacts are not always private. These contacts are divided into procedural contacts, work-related contacts, and behavioral or disciplinary contacts. They are also separately coded according to whether they are initiated by the teacher (teacher-afforded) or by the child (child-created). The coding also reflects certain aspects of the teacher's behavior in such contacts. *Work-related* contacts include those teacher-child contacts that have to do with the child's completion of seat work or homework assignments. *Procedural* contacts include all dyadic teacher-child interaction which is not coded as work-related contacts or as behavioral contacts. *Behavioral* contacts are coded whenever the teacher makes some comment on the child's classroom behavior; work-related or procedural matters are not involved.

Supplementary Observations

In addition to recording teacher-child interactions, the observer should also

record the presence or absence of resources in the classroom as well as impressions of the classroom climate.[3] We are not only interested in the types of resources available (such as nonteaching staff, equipment, and reference books), but also the use that is made of these resources. We are interested in visual or other clues which would indicate that teacher and students are promoting interracial contact and friendships. These supplementary observation processes could be recorded during the periods when the teacher is lecturing or interacting with a student who is not being observed.

In order to get a feel for the climate of the school in general, the observer should follow the classroom to the playground during recess and the school cafeteria during lunch. Again, the frequency and type of interracial contact among the students should be recorded.[4] At the conclusion of the day, the observer would record overall impressions of the school and the classroom observed. These impressions would be largely judgmental, but would serve to suggest how successful this desegregated school has been along certain key dimensions.[5]

Open Classrooms

Essentially the same observation procedure can be used in open classrooms. The one exception to this procedure concerns the number of students per classroom to be observed. Two students should be observed in an open classroom. At fixed intervals, the observer would alternate between these two students, recording the teacher-child interactions as described above, as well as any interactions between the students and other adult figures in the classroom. The observer could also conduct mini-interviews with students periodically to determine their reactions to classroom activities.

Data Analysis

Various analyses can be performed with this data base. The relative amount of time spent in different activities or the types of interactions which tend to occur in different types of activities can be calculated. Comparisons can be made of the *quality* of teacher-child interaction in different individuals and groups, despite differences in *quantity* of dyadic interactions with the teacher. Comparisons can be made concerning the use of affirmation or praise for correct answers and, conversely, the use of negation and criticism for incorrect answers. The data can be analyzed to determine if teachers sustain interaction longer for some students than for others. From this proposed study it could be determined whether or not teachers tend to treat students differently according to ethnicity or achievement level. This study will serve, as well, to validate and expand upon our findings elsewhere concerning the effects of school desegregation.

[3] See the Appendix for suggestions of types of information to be recorded on supplementary form.

[4] See Appendix for suggested types of information to be collected.

[5] See Appendix for form. For an example of how an overall assessment of teachers can be used, the reader is referred to Nancy St. John (1971).

10. SAMPLING

SAMPLE OF SCHOOL DISTRICTS

The target universe will be those school districts in the continental United States with sufficient numbers of target minorities for desegregation to be a potential issue. Operationally, we have chosen to translate "sufficient numbers" into "3000 population" according to U.S. 1970 census; the criterion might have been defined in terms of enrolled students, but data for this variable would be more difficult to obtain. The target minorities are blacks, Mexican-Americans, and Puerto Ricans. It will be assumed that there are no districts with sufficient concentrations of Mexican-Americans or Puerto Ricans in Census Region V (southeast: Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia). For the remaining states, 1970 census counts of "persons of Spanish language"[1] will be used to determine whether a district qualifies on the basis of its Mexican-American or Puerto Rican population. Adopting these criteria, an examination of 1970 census data aggregated at school district level[2] reveals a sampling universe numbering 1399 districts distributed as follows:

Black population \ Spanish-language population	<3,000	3000-<25,000	25,000 and over
<3,000		341	17
3000-<25,000	830	83	15
25,000 and over	43	49	21

The sampling of districts will be accomplished in four separate samples (the first three will be for districts that qualify on the basis of their black populations):

1. *The North Sample.* One hundred districts will be sampled from a sampling universe of approximately 400 districts that qualify on the basis of their black populations. The universe includes all CONUS states except Texas and the 12 states in Census Region V.
2. *The South/Coleman Sample.* Forty districts will be selected from a sampling universe comprised of those *non-Metropolitan* counties in Region V and Texas for which all high schools were included (and responded) in the Coleman study (that is, those counties for which the within-county sampling fraction was 1.0).

[1] "Persons of Spanish language" is a term applied by the Bureau of the Census to persons who reside in a household where Spanish was spoken in the childhood home of the head-of-household or his spouse.

[2] Census data aggregated by school district are available in the 1970 Census Fourth Count (Population) School District Data Tapes, distributed by the National Center for Educational Statistics. These computer tapes were prepared by Applied Urbanetics, Inc. under contract to USOE.

Excluded from this sampling population would be all districts with fewer than 3000 black persons in residence.

3. *The South/non-Coleman Sample.* Sixty districts will be selected from the universe of districts in Region V and Texas that have 3000 or more black residents, but that are not included in the sampling universe for the South/Coleman sample.
4. *The Auxiliary Spanish Sample.* It has been noted in Sec. 6 that a large number of districts in the North sample are expected to have significant Spanish-language populations. These districts will be supplemented by a sampling of 20 districts from among those that contain more than 3000 Spanish language persons but fewer than 3000 black persons. In this case, the sampling population numbers 357 districts, and includes all states outside Region V.

Each of the four samples will be drawn in two phases. In each case the first phase is a screener sample of a relatively large number of districts. Data will be gathered on districts in the screener samples by methods that have been detailed in Sec. 6. Some of these data will be used to accomplish a more refined stratification of the screener samples. The stratified screener samples will then serve as sampling universes for the second phase, or ultimate, samples.

Somewhat more than half the districts in each of the four samples will be sampled in matched pairs. The remaining will be selected in "antimatched" pairs; that is, pairs that are similar as possible in all respects, except that one district is segregated and the other is highly desegregated. The antimatches have the specific purpose in the subsequent sampling of high schools of providing clusters in which desegregated schools are selected from desegregated districts and segregated schools are selected from segregated districts, thus avoiding a measure of self-selection bias attributable to the school populations.

Control Variables

We are recommending that sampling be highly controlled to assure that samples accommodate to the needs of various analyses in the overall study. The most important variables will be cited briefly here; their logic with respect to analysis has been explained in Sec. 6. It should be noted that all of these variables are not expected to be used in each of the four samples.

There are five variables for which data can be assembled from centrally located sources. These variables will be utilized in both sampling phases.

District Population. 1970 counts of school district population are readily available from computer tapes distributed by the Bureau of the Census. Distribution of *all* districts (regardless of racial-ethnic composition) in the United States by size is given below:

$< 10{,}000$	$10{,}000 - < 25{,}000$	$25{,}000 - < 50{,}000$	$\geq 50{,}000$
8082	2628	993	718

Ethnicity. Here we are concerned with the proportions of school district populations that are black or Spanish language. It would be useful to combine the two into a single categorization, e.g., Low Black and Low Spanish, Low Black and Medium Spanish, Low Black and High Spanish, etc. The class limits separating Low from "Medium" and "Medium" from "High" should be chosen to divide the universe into strata that are reasonably comparable in size and that have analytic relevance, that is, relevance with respect to characteristics of desegregation.

Socioeconomic Status. SES can be measured by the median educational attainment, or percentage of adults who are high school graduates.

Degree of Urbanness. In some cases, it may be useful to stratify between more and less urban districts. This stratification might be on the basis of the census variable, "percent urban," or whether districts are within or outside SMSAs. Neither approach is without fault, but either would reasonably assure that both urban and rural areas are represented in the sample.

Geographic Area. The Northern Sample is divided into three regions: The three Pacific coast states plus Nevada, Pennsylvania and all states to its North and East, and the remaining area, from the Midwest to the Rocky Mountains. The South is divided into two areas: one is the Central South (Louisiana, Mississippi, Alabama, Georgia, and South Carolina); the other will be called the Peripheral South.

At least three additional variables will be made available through the screener surveys. These can be used to effect more refined control for the second phase samplings.

Degree of School Segregation. It is recommended that an adaptation of the index of dissimilarity, sometimes called the Taeuber Index, (Taeuber and Taeuber 1965, 1972), be used to indicate the degree of segregation within a school district. This is an index that was originally used to measure residential segregation, but which has also proved useful in measuring school integration (Rossell and Crain 1973).

Degree of Student Reassignment. In Sec. 6, five levels were suggested for this variable: no desegregation action taken, desegregation without reassignment, desegregation with little reassignment, desegregation with medium reassignment, desegregation with much reassignment.

Whether Desegregation Is Court-Ordered. In the North there would be two levels: court-ordered and not court-ordered.

As previously noted, the variables described are candidates for controlling the four district samplings; all are not appropriate for each sample. In selecting from among these variables, the sampler must satisfy himself of their logic for the particular sample being drawn.

When a set of control variables has been selected, their distribution over the sampling universe should be examined to see if a smaller set might provide the same degree of sampling control. Some variables may be redundant. For example, initial investigation suggests that the nationwide correlation between the Taeuber index and counts of students who are reassigned through desegregation is about 0.7. This correlation is not high enough to dismiss one of the variables out-of-hand, but it suggests the possibility that the correlation might be higher in one or more of the four sampling populations.

Even if redundancy is not present, there may be opportunities to combine two variables into one. Such combinations might make the mechanics of selection simpler. For example, if one variable has only two levels and another has three, it would probably be more convenient to combine them into a single stratification with six levels (2 x 3) than to deal with them separately.

One issue to be dealt with in defining the levels for each variable in the four sampling frames is whether class boundaries for levels ought to be determined independently for each frame, or whether class boundaries ought to be the same for all frames. Since districts may distribute over the variables differently for different sampling universes, independent specification of class boundaries may produce strata that are more meaningful from the standpoint of sampling control. On the other hand, it may make more sense for data display and intersample analyses if the class boundaries are uniform for all samples. One approach is to designate a set of uniform boundaries for all four samples, then designate additional class boundaries as needed for each of the four samples; this would make it convenient for the analyst

to collapse data from each sample into a uniform set of strata. Another approach would be to designate class boundaries independently for purposes of sample selection, then simply post-stratify for purposes of analysis; this procedure may be risky where the distributions of sampling populations across a given variable are disparate.

Sampling Techniques for Utilizing Multiple Control Variables

One approach to designing a sample for which several stratification variables or factors are intended is to simply array the universe units among the cells of a multidimensioned matrix, with one dimension for each factor; then sample independently from each cell of the matrix. For example, in order to stratify districts by degree of desegregation at three levels, SES at four levels, and race-ethnicity at four levels, one might first allocate the universe of districts among the 48 cells of the corresponding 3 x 4 x 4 matrix. The number of districts sampled from each cell might be set proportional to cell size (i.e. number of districts within each respective cell) if it is desired to give equal selection probabilities to all cells, or some cells might be oversampled in order to provide adequate sample sizes for particular analyses or to accommodate uneven distributions of districts across cells. From the standpoint of sample selection, such a scheme is equivalent to simple one-way stratification where strata are defined by multiple criteria.

In this section, this sort of design will be called *factorial sampling* to emphasize its structural relationship to factorial experimental design; this relationship is of interest because, by virtue of the fact that factorial design provides observations of all combinations of the various levels of all factors, it is possible to isolate interactive effects as well as the main effects of the factors themselves. The other approaches to sample design that will be discussed do not guarantee this property.

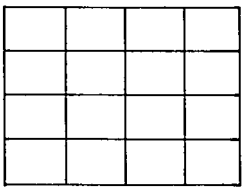
Although factorial sampling appropriately provides for the analytic needs of surveys, it has disadvantages in its application. With the sample of districts for the desegregation study, for example, the sample sizes are small relative to the number of cells, and the distribution of districts across these cells is uneven at best. For sample selection to be probabilistic (a necessary condition for unbiased estimation of population parameters), a considerable amount of weighting in the estimators would be required.

While the analytic purposes of the desegregation studies are undoubtedly more important than the purely descriptive (i.e., population parameter estimation), the descriptive properties can be maintained without great damage to analytic properties if consideration is given to multi-stratification designs that do not require selections from each cell. As will be noted, these designs provide an especially attractive way to incorporate multiple stratification control, since many cells in the sampling universes may be empty of districts, and since in some cases the number of cells may exceed the sample size.

Since these methods may be relatively unfamiliar, we provide a brief background on the development of techniques beyond simple stratification for utilizing auxiliary variables for sampling control. Particular attention is focussed on the controlled sampling of Goodman and Kish (1950) and probability lattice sampling, as these schemes will later be recommended (along with factorial sampling) for various stages of the four samples of school districts.

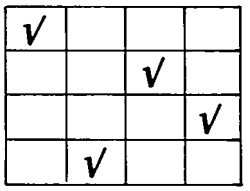
Since 1940, a number of approaches have been documented that impose controls beyond simple stratification without sacrificing the integrity of probability sampling. The unifying characteristics of these schemes can be illustrated by depicting a case where maximum stratification is imposed for each of two factors; for $n = 4$

this means that the population is partitioned into four groups for each factor. The population may be represented by a 4x4 matrix, and for multi-stage designs the cells become primary sampling units:



The idea is to randomly select four of the 16 cells such that each row and each column is represented. Any grouping of cells that satisfies this requirement will be referred to as an eligible pattern.

Lattice Sampling. The term lattice sampling was applied by Yates (1953) to a class of sample designs that began surfacing in the late thirties, apparently inspired by Fisher's (1935) descriptions of the Latin square and Graeco-Latin square techniques in experimental design. For the two-way controlled sampling example described above, lattice sampling essentially involves selecting four of the sixteen cells with equal probabilities in such a manner that all strata in both dimensions are represented. One such resulting pattern might be:



The method is easily modified to accommodate three stratification criteria.

Lattice sampling has one serious drawback for multi-stage surveys in that cells are necessarily drawn with equal probabilities. If it is desired to preserve equal probabilities for the ultimate sampling units, then each cell must contain an identical number of them. Unfortunately, naturally occurring populations rarely can be forced into such multi-dimensional symmetry. More typically, multi-stratification would result in varying cell sizes (or MOS), such as the simple two-dimensional case below:

.0	.2	.5	.3	1
.2	.2	.2	.4	1
.3	.4	.1	.2	1
.5	.2	.2	.1	1
1	1	1	1	4

where numbers within cells depict the relative cell sizes, numbers on the margin are the totals for the cells in the respective strata, and the matrix total is in the lower right-hand corner; the matrix total is set equal to the sample size for convenience. When simple lattice sampling is applied to such a frame, it is necessary to incorporate weighting in the estimators.

Frankel and Stock (1942) are generally credited with introducing this technique under the label of Latin square sampling. The following year, Tepping, Hurwitz, and Deming (1943) examined the efficiency of several versions of a related technique, *deep stratification,* in which the symmetry of the Latin square matrix was distorted as required to provide cells of equal size. For our two-dimension example, the idea is to first divide the population into four equal-sized groups along one dimension, then divide each of the four groups into four equal-sized sub-groups along the other dimension. Although the 16 cells are now of equal size, the class limits for the sub-groups vary for different primary control categories, which may be undesirable when separate estimates or analyses are to be prepared for different levels of the second control variable. The scheme may also be faulted for inefficiency, since the control offered by the second factor (i.e., the column effect) is compromised.

Yates (1953, 10.14) provides a brief discussion of Latin square sampling and deep stratification, and suggests the more general term *lattice sampling.* Hansen, Hurwitz, and Madow (1953, 11.4) describe Latin square sampling, and present an expression for sampling variance which they attribute to Jerome Cornfield and W. Duane Evans (no citation given). Variance and the estimation of variance for four approaches to lattice sampling are the main topics of a paper by Patterson (1954).

In a much later paper, Bryant, Hartley, and Jessen (1960) present an approach for sampling from a population classified in two dimensions where both cells and marginal totals for strata may be unequal, although cell sizes must be nonzero. The method, termed *two-way stratification,* essentially involves converting the matrix to a symmetric square so that selection can proceed as with lattice sampling.

Controlled Selection. The lattice sampling techniques described above exhibit problems in implementing the idea of unequal probability selection for a population stratified in more than one dimension. Goodman and Kish (1950) have described a technique whereby formal stratification is obtained on one dimension, and other dimensions are accommodated informally by placing restrictions on the selection within strata. In this manner, a balance is maintained with respect to other stratifying variables. They label this general approach *controlled selection.* The distinguishing feature seems to be that with controlled selection some eligible patterns are excluded from consideration, either purposively or arbitrarily. Some patterns may be preferred over others, and this preference is manifested by interfering with the independence of selection within strata.

In particular Goodman and Kish describe the following sort of scheme. Returning to the example population of the preceding pages, the formal control (i.e., conventional simple stratification) is indicated by Roman numerals, and Arabic numerals label the strata for the informal control:

	1	2	3	
I	.5	.0	.5	1.0
II	.5	.3	.2	1.0
III	.3	.3	.4	1.0
IV	.3	.4	.3	1.0
	1.6	1.0	1.4	

The objective is to designate a feasible set of preferred samples, or patterns. Each pattern will conform exactly to the row margins and approximately to the column margins; that is, it must contain one cell from each row, and be roughly representative of the three columns. In addition, each pattern is assigned a selection probability such that (1) the sum of selection probabilities for all patterns in the feasible set is equal to 1.0, and (2) the sum of selection probabilities for those patterns which contain a given cell is proportional to the cell size. Thus, by observing the selection probabilities in selecting one of the patterns, the probability for any cell is proportional to its size. A bookkeeping worksheet suggested by Goodman and Kish provides for the orderly designation of the feasible set.

Whereas Latin square sampling, for example, admits all eligible patterns, any given application of controlled selection typically admits only a subset of the total population of eligible patterns. Since the designation of this subset is purposive, depending on the whim of the sampler, the set of joint selection probabilities is not unique, so there is no unique sampling variance for a given sampling frame (in contrast to lattice sampling or simple stratified sampling). There is a unique variance for a given feasible set, but it typically cannot be estimated unbiasedly because there are often pairs of cells with zero joint probabilities, even when more than one cell is designated from each row and column. Patterson (1954) suggests drawing more than one pattern from the set, then basing variance estimates on contrasts between the selected patterns; this approach generally detracts from sampling efficiency.

The selection of cells is nevertheless with probabilities exactly proportional to cell size, so no bias is introduced in the estimation of averages and totals, and the self-weighting samples suggested by Hansen and Hurwitz (1943) are obtainable.

Goodman (1963) describes an approach for estimating sampling variance unbiasedly that requires the selection of additional units exclusively for that purpose.

For some universes, a perfect feasible set, in which all patterns conform to all controls, is unobtainable if the restrictions on selection are too stringent. In view of this possibility, Jabine (1966) has suggested a strategy for developing a test that would indicate whether this is the case for a given universe.

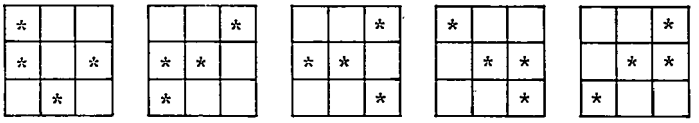
Probability Lattice Sampling. Hess and Srikantan (1966) suggest a procedure for designating feasible patterns that conform exactly to integer margins with two or three factors rather than just one; for the three-dimensional case, they also suggest nested contraints (i.e., controlling exactly by the joint margins of each pair of factors). They note that in many cases a completely perfect feasible set cannot be obtained, but they suggest a format for designating patterns such that the number of imperfect patterns is minimized.

Jessen (1970) adapts his Method 2 from an earlier paper to the problem of sampling from multi-stratified universes. The adaption is very similar to the development of Hess and Srikantan, except that fewer restrictions on the feasibility of patterns (e.g., patterns need not satisfy nested constraints when sampling from universes with more than two dimensions) greatly diminishes the problem of whether a perfect set can be designated from a given universe, e.g., empty cells present no problem. Method 2 also incorporates a strategy for minimizing the number of patterns required to complete a feasible set.

Consider the problem of sampling four cells from a matrix with the following distribution of MOS:

.5	.0	.5	1.0
.8	.6	.6	2.0
.3	.4	.3	1.0
1.6	1.0	1.4	4.0

a feasible set of patterns and the associated selection probablities might be:



.4 .2 .2 .1 .1

Note that the sum of selection probabilities of all patterns that contain any given cell is equal to the designated probability for that cell, and that the expectation over all patterns of the number of cells chosen from any row or column is equal to the respective marginal restriction. The mechanics of Method 2 are explained in detail in Appendix D.

In a subsequent paper, Jessen (1973) labels this general approach to satisfying marginal and cell constraints as *probability lattice sampling,* apparently to denote its relationship to lattice sampling; both methods satisfy marginal and cell constraints, but lattice sampling as described by Yates and by Patterson is an equal probability scheme. The same article develops an approximate model of the variance structure for universe with unequal cell sizes, and provides a simple comparison of three approaches for estimating sampling variances for probability lattice sampling.

Sample Selection

The selection schemes for the four samples will be complex, given the excessive degree of sampling control that is desired and the uncertain distribution of the sampling universe across the control variables. The schemes described in the following pages give our sense of how sampling should or could proceed, but they have been formulated without the benefit of a thorough description of the universe. After examining the universe, the contractor may decide that certain variations to our approach are more appropriate. Our main concerns, in order of importance, are that the sampling schemes (1) provide representation over the various variables that is adequate for analysis purposes, (2) provide for selection that is probabilistic (i.e., probabilities are known, although not necessarily equal), and (3) maintain as much simplicity as is consistent with the other constraints. This will enhance prospects for developing population estimates and approximations for sampling variance.

The numbers specified in the descriptions of sampling schemes are target sample sizes and should be inflated to anticipate nonresponse. For the first-phase (screener) samples, the inflation factor should be 10 or 15 percent. The factor might be as high as 35 percent for the second-phase sampling, since there may be a strong reluctance on the part of districts to participate in the study. The contractor should try to get a fix on the magnitude of this reluctance during the screener surveys. In any case,

considerable attention should be given to the problem of minimizing district refusal; this is especially important given the large number of categories for which representation is desired.

In the selection schemes that follow, oversampling is indicated twice. It may be necessary to oversample more often if the numbers of districts with certain desired characteristics are very sparse; for example, the number of northern school districts experiencing court-ordered desegregation. If much oversampling is required, it may be difficult to adhere to the multiple stratification that is desired. In such cases, it would be necessary to rejudge the relative merits of oversampling versus retaining the unbiased, multi-controlled sampling structure.

Although probability lattice sampling (PLS) is suggested more often than factorial sampling for achieving multi-controlled samples, factorial sampling is easier to implement and should be investigated as a possibility wherever both universe size and sample size are large relative to the number of cells in the sampling frame.

Because of the complicated nature of the sampling schemes, it may not be possible in some cases to obtain a complete feasible set in which all designated patterns conform exactly to the margin constraints for rows, columns, etc. A practical criterion would be to accept an incomplete feasible set if the sum of selection probabilities for balanced patterns total 0.85 or 0.90. Selection could be restricted to those patterns without greatly distorting unbiasedness.[3] If the sum for balanced patterns is much less than this criterion, alternatives for simplification should be considered (such as fewer factors, fewer levels per factor, or possibly relaxing the requirement for probabilistic selection).

The North Sample. There are fewer than 400 school districts in the northern states where the black population exceeds 3000. For this reason, we recommend that the first-phase survey be a canvass of all qualifying districts rather than a sample.

For the second-phase sampling, selection can proceed in two stages. In the first stage, PLS is employed to select, say, fifty cells from a five-dimensioned sampling frame stratified according to the Taeuber index, race-ethnicity, SES, geographic region, and district population. In the second stage, two districts will be selected from each cell, utilizing controlled sampling to guarantee either similarities or differences on various levels of administrative desegregation (degree of student reassignment) and court involvement.

To construct the frame for the PLS, the sampling universe of 400 or so districts should first be stratified according to the Taeuber index into three levels. The 15 percent of the sampling universe (approximately 50 districts) with the lowest index values will comprise the "most highly desegregated" stratum. The 45 percent of the universe with the next-lowest index values will form the "desegregated" stratum. The remaining 40 percent are the "segregated" districts. This group is further divided into two strata: the 25 percent of segregated districts (10 percent of the universe) that are most likely to desegregate (as predicted by the procedure described in Sec. 6) and all other segregated districts.

The target sample sizes in terms of the Taeuber index stratification are:

30—most highly desegregated
30—desegregated
20—segregated, most likely to desegregate
20—other segregated

[3] Evidence that such a procedure, though slightly biased, typically results in smaller mean square errors is reported in Sumner (1973). Although Sumner's study deals with a household sampling universe, there is no reason to suspect the phenomenon to be household-specific.

Comparing these sample sizes with stratum sizes, it is apparent that the sampling rates in the first and third strata are three times higher than those in the second and fourth strata. Before stratifying by the other variables, selection probabilities should be adjusted to provide for this oversampling. The adjustment can be accomplished by assigning districts in these strata a measure-of-size that is three times that of districts in the other two strata. Thus, if the sampling universe numbers 400, selection probabilities for districts in the two oversampled strata will be one-half, whereas selection probabilities for districts in the other strata will be one-sixth.

The race-ethnicity factor should have enough levels to adequately control for different concentrations of black and Spanish-language populations. It seems likely that six levels (three for black times two for Spanish) would be most appropriate, although it may turn out better to use as few as four or as many as nine.

It should be possible to combine the SES, the district size, and the geographic region variable into a single factor. With three levels for SES, two levels for district size, and three levels for area, the number of levels for the combined factor would be eighteen (3 x 2 x 3).

Together, the three factors partition the sampling universe into a 3 x 6 x 18 matrix, with one dimension for each factor. This matrix will provide the sampling frame. The mechanics of PLS will be greatly simplified if the class boundaries of the levels are set so that the expected number of cells to be chosen from each level (stratum) is an integer.

One hundred school districts will be sampled from this frame, 52 in *matched* pairs and 48 in "*antimatched*" pairs. As was described earlier, the antimatches contain one district that is highly desegregated and one district that is segregated; they are matched with respect to the other sampling criteria.

The simplest way to control the numbers of each kind of pair is probably to sample them separately. We will suggest a somewhat more complicated method that is designed to maintain more nearly equal selection probabilities for all districts. Briefly, PLS will be employed to make 100 cell selections[4] from the sampling frame, then one district will be selected at random from each sampled cell. In designating each pattern of cells, cells will first be designated from which the antimatched district pairs will be selected; cells for the matched district pairs will then be designated. This procedure amounts to overlaying PLS with the more informal controlled sampling of Goodman and Kish (1950). The informal control in this case guarantees appropriate numbers of cells for drawing 24 antimatches and 26 matches.

Before proceeding with selection, it will be necessary to assure that each sampling unit in the district sample contains at least two high schools in order to accommodate the subsequent sampling of high school clusters from the sample districts. Each district containing one or no high schools should be combined with another district so that each sampling unit contains at least two high schools. If a district has no high school (i.e., elementary schools only), it should be combined with a randomly chosen district within the same cell of the sampling frame; if there are no other districts in the same cell, a district should be chosen from the district-occupied cell in the same Taeuber index level and the same race-ethnicity level that has the most similar SES-size-region characteristics.

For the first stage sampling of 100 cells, PLS is employed. In designating each feasible sample, cells from which antimatched district pairs can be selected are

[4] The 100 cells will not be distinct. Cells from which matched district pairs are desired will be selected twice. The probably uneven distribution of districts among cells will require additional multiple selections of cells containing many districts. It is not that a single cell may be targeted for a matched pair as well as one-half of an antimatched pair. The implication for the selection of districts within cells is that the number of districts selected from each cell must equal the number of times that cell was selected.

designated first. The cells are designated in pairs, pairing occupied cells in the "highly desegregated" stratum with cells in the "segregated" strata (the "segregated, about to desegregate" stratum and the "other segregated" stratum); both cells in each pair must belong to the same race-ethnicity level and same SES-size-region level. When 24 such pairs have been designated, 26 additional pairs are designated from among all occupied cells in the matrix, but in this case each "pair" is actually the same cell designated twice;[5] a matched district pair will eventually be sampled from each of the 26 cell "pairs." In each pattern thus designated, the 50 cell pairs taken together should satisfy the marginal constraints (i.e., sampling quotas for rows, columns, etc.) of the sampling frame.

After one of the feasible patterns has been selected in accordance with the PLS procedure, district pairs will be selected. Antimatched district pairs come from the 24 cell pairs so targeted; the district pairs should be determined by randomizing from among the eligible alternatives in each cell pair. The selection of matched district pairs within the targeted 26 cells should not be independent, but should be done in such a fashion that not only are two districts chosen from each cell, but the sample of 52 districts includes at least five to ten districts from each level of the following control variables:

Administrative Desegregation
(1) no desegregation
(2) desegregation, no student reassignment
(3) desegregation, little reassignment
(4) desegregation, medium reassignment
(5) desegregation, high reassignment

Court Involvement
(1) court ordered
(2) not court ordered

An appropriate scheme for obtaining representation over these variables could be modeled after the Goodman and Kish controlled sampling that was described earlier.

The 100 districts in the 50 pairs should be unique, even though this was not necessarily so for the 100 cells. For example, a single district should not be part of an antimatched pair as well as a matched pair.

The South/Coleman Sample. The sampling population for the first phase (screener) of the South/Coleman sample will be composed of those rural districts in Census Region V and Texas that were included in the Coleman study and which are located in counties in which the Coleman study canvassed all high schools (i.e., those counties for which the within-county sampling fraction was 1.0). Nationwide, the Coleman study included 821 high schools in counties outside Metropolitan Areas and about half of these were from counties where the sampling fraction was near 1.0. In Region V, 178 high schools were included from counties outside Metropolitan Areas, hence it might be supposed that the number from counties where the sampling fraction was near 1.0 is fewer than 100. Adding Texas, it seems reasonable that the total number of *districts* eligible for the South/Coleman sample might be approximately 100. The identity of these districts is not public information, hence it will be necessary for the contractor to work through the appropriate government

[5] The exception to this rule is for cells that contain only one district. In such cases, the cell should be paired with another nearby cell. Such cells can be designated once, if paired with the single designation of another nearby cell. The nearby cell should be on the same level with respect to the Taeuber and race-ethnicity factors, and similar with respect to its SES-size-region characteristics.

agencies to identify the sampling universe. Recall that it is necessary that the black population in the districts number at least 3000. The first phase (screener) survey should be a canvass of all 100 or so districts that fit this qualification.

As with the North sample, second phase sampling can proceed in two stages. In the first stage, PLS can again be employed to select 40 cells (16 targeted for antimatched district pairs and 24 targeted for matched pairs) from a multi-stratified sampling frame. In this case, it is probably sufficient that only four factors be used: the Taeuber index, race-ethnicity, SES, and geographic region. In the second stage, one district will be selected from each sampled cell, utilizing controlled sampling to guarantee various levels of student reassignment.

Preparation of the sampling frame for PLS will be simpler than before, since there is no need for oversampling. Three levels are sufficient for the Taeuber index: high desegregation, medium desegregation, and "token" desegregation. The number of levels for race-ethnicity and the SES factors should be at least three. The two-region strata (Central South and Peripheral South) have already been described. The SES and region factors should be combined to a single factor for sampling convenience. The resulting sampling frame would be a 3 x 3 x 6 matrix.

As before, class boundaries for the levels should be set so that the expected number of cells to be chosen from each level is an integer (or approximately integer in the case of the Region factor). Selection will be even more simplified if all strata for a given factor contain equal numbers of districts as well as integer margin expectations. This would require altering the matrix dimensions so that the number of levels for each factor is divisible into 40. For example, a 4 x 4 x 8 (4 levels for Taeuber, 4 levels for race-ethnicity, 8 levels for SES-Region) might be employed; after sampling is completed, the matrix could be collapsed into 3 x 3 x 6 for analysis purposes.

Adjustments for districts with one or no high schools, selection of cells targeted for mismatched and matched pairs, and the selection of districts within cells should proceed as with the North sample. The within-cell selection of districts for matched pairs should be accomplished in such a fashion that the 24 districts include four or so districts from each level of (1) no student reassignment, (2) little student reassignment, (3) medium student reassignment, (4) high student reassignment. Again, the Goodman and Kish controlled sampling is recommended for this purpose.

The South/Non-Coleman Sample. An additional 60 districts will be selected from the universe of districts in Region V and Texas that have 3000 or more black residents, but which are not included in the sampling universe for the South/Coleman sample. The number of districts that qualify is probably in excess of 600.

The first phase survey should be a sampling of 200 of these districts in order to keep data handling to manageable size. The sampling should be controlled by race-ethnicity (three levels), SES (three levels), district size (two levels), and geographic region (two levels). Given the large universe and sample sizes (600 and 200, respectively), and the expectation that the number of levels per factor can be kept at three or below, it may be reasonable to utilize factorial sampling. Accordingly, the universe would be stratified four ways into a 3 x 3 x 2 x 2 matrix containing 36 cells. Each cell would be treated as a separate stratum, with independent but proportional sampling within each. This approach would result in a screener sample that reflects the joint distribution of the sampling universe across the four variables. On the other hand, if more levels per factor are required, it may be more appropriate to resort to PLS, in which case the screener survey would reflect the simple distributions of the universe across each of the four variables, but not necessarily the joint distribution. With PLS, the selection of districts within cells would be accomplished with independent simple random sampling. Regardless of whether PLS or factorial

selection is employed, adjustments should first be made for districts with one or no high schools following the same procedure as for the North sample.

The second phase sampling proceeds as with the North sample, with some simplification. Sixty districts (12 antimatched pairs and 18 matched pairs) are selected in two stages from the 200 sampled in the screener survey. In the first stage, PLS is employed to select 60 cells from a sampling frame that is stratified by the Taeuber index, race-ethnicity, SES, district size, and geographic region. In the second stage, one district is selected for each cell selection for a total of 30 pairs. An informal control should be utilized for the 18 matched pairs to guarantee representation of various levels of student reassignment.

In preparing the sampling frame for PLS, it would be convenient to combine two of the factors (e.g., region and district size) as was suggested for the North sample. It would also be convenient to set class boundaries that equalize the numbers of districts in each stratum. After sample selection, the districts could be recategorized for analysis purposes.

The Auxiliary Spanish Sample. An additional 20 districts will be sampled from among the 357 districts in the nation (outside Census Region V) which contain more than 3000 persons of Spanish language but fewer than 3000 black persons (and therefore were excluded from the other three sampling universes). Sampling procedures should be similar to those used for the South/non-Coleman sample.

A first phase (screener) survey will be administered to a sample of 80 districts. Either factorial sampling or PLS might be employed. The sampling frame should be stratified by SES (three levels), population size (two levels), geographic region (with Texas, New Mexico, and Arizona as one category, and the remainder of the non-South divided as before, for a total of four levels), and degree of urbanness; oversampling should be employed if necessary to assure that at least 15 of the districts are urban.

For the second phase, a stratification based on the Taeuber index (three levels) should be added to the other four factors. As before, factors can be combined and numbers of levels modified as required for sampling convenience. PLS should then be used to select 20 districts, with four antimatched pairs and six matched pairs.

SAMPLE OF SCHOOL CLUSTERS

High School Clusters

As was explained in Sec. 6, clusters of four or five high schools will be selected from among the high schools in the districts that have been sampled. The four-school clusters consist of two mixed[6] schools matched with respect to Anglo SES[7] and dominant minority SES, one Anglo school that matches the mixed schools on Anglo SES, and one minority school that matches the mixed schools on dominant minority SES. If the mixed schools contain both black and Spanish heritage populations in significant proportions (over 10 percent), there will be two minority schools in the cluster, one matching each of the minority populations on SES.

[6] A mixed, or desegregated, school is defined here as containing an Anglo student body of at least 25 percent and at least one minority group that is at least 10 percent of the student enrollment.

[7] This is the SES of residents of the attendance area. Regardless of the SES surrogate variable that is used to represent SES, it may be difficult to obtain the information aggregated at attendance area level. Schemes for approximating this variable should be investigated, such as using the SES measurement for the census tract where the school is located.

In the case of the antimatched district pairs, the mixed schools will be selected from the desegregated district, and the segregated schools will be selected from the segregated district. Two such clusters may be selected from each antimatched pair. In each district sampled as part of a matched pair, a complete cluster of high schools may be selected.

With 60 antimatched pairs and 80 matched pairs in the sample, there is a potential for 280 high school clusters. The target sample size is 130. The sample would be distributed among the four district samples as shown in Table 10.1. Table 6.4 includes numbers that indicate our expectation of how the clusters will distribute across ethnic groups.

The method adopted for selecting clusters should accomplish two purposes. It should provide that clusters be selected from those districts where the best matches are possible, and it should provide for random selection of clusters within districts with two or more equally well-matched clusters. The following approach is suggested.

Sampling from Antimatched Pairs. For the antimatched district pairs, all high schools in the desegregated districts and black high schools in the segregated districts are stratified into four levels according to black SES; the strata should each contain about the same number of schools. Similarly, all high schools in the desegregated districts and dominantly Anglo high schools in the segregated districts are stratified into four levels according to Anglo SES. Finally, all high schools in the desegregated districts and dominantly Spanish heritage schools in the segregated districts are stratified into four levels according to SES of the Spanish heritage populations. Thus, all high schools in the desegregated districts have been stratified into a 4 x 4 x 4 matrix, and each school in the segregated districts has been stratified into one of the 1 x 4 matrices.

Next, up to two quartets are itemized for each district pair. In each quartet, there must be two schools from the desegregated district that match (i.e., belong to the same stratum) with respect to Anglo SES and the SES of the dominant minority, an Anglo school from the segregated district that matches the desegregated schools in Anglo SES, and a minority school from the segregated district that matches the desegregated schools in the SES of the dominant minority. If there is a second minority population (black or Spanish heritage) of significant proportion in the desegregated schools, and if there exists in the segregated district a high school of the same minority with matching SES, that high school should also be included in the cluster; thus, in such cases, the cluster contains five high schools.

Table 10.1

DISTRIBUTION OF SAMPLE AMONG DISTRICTS

	North Sample	South/ Coleman Sample	South/ Non-Coleman Sample	Spanish Auxiliary Sample
Number of antimatched pairs	24	12	16	8
Potential number of clusters	48	24	32	16
Target sample	32	6	15	4
Number of matched pairs	26	18	24	12
Potential number of clusters	52	36	48	24
Target sample	38	9	20	6

If more than two clusters are obtainable from a given district pair, the number should be reduced to two by randomization. Randomization should also be employed if, for a given cluster, alternative schools are available for a given cluster position.

After the clusters have been identified from the antimatched district pairs, the clusters should be listed in four groups, each group corresponding to one of the four district samples. Within each group, the clusters should be listed in some systematic order that relates to the structure of the district sampling frame. For example, clusters from the North sample might be listed according to SES-size-region factor within the race-ethnicity factor. An appropriate sampling interval should then be determined, and a sample of clusters selected using systematic sampling. From the North sample, for example, the target sample size from the antimatches is 32; if the total clusters itemized from the North sample were 48, the appropriate sampling interval would be $48/32 = 1.5$.

If the total clusters itemized from any of the four district samples is less than the target sample size, all clusters should be included in the sample. The deficit in the sample should be filled by recategorizing high schools from pairs that did not yield two clusters, this time using three levels of SES for each population group rather than four. The itemization and sampling of clusters should be repeated as above, using the deficit rather than the target sample size to determine the sampling interval.

Sampling from Matched Pairs. The same method should be used to sample clusters from the matched district pairs, except that all schools in a given cluster come from the same school district. In this case, segregated schools are defined as those that are either less than 25 percent Anglo or less than 10 percent minority.

Some districts may not contain four high schools; such a district should be randomly assigned to another district in the same or adjacent cell of the district sampling frame.

The systematic sampling procedure described above for antimatched pairs should be modified so as not to preclude the possibility of including clusters from both districts of the matched pair in the sample. One approach would be to sample separately from among those where the district pairs yielded two clusters.

Elementary and Junior High Clusters

A junior high school (or middle school) cluster and an elementary school cluster will be selected from among all the feeder schools within the combined attendance areas of each high school cluster. The sampling procedure can be similar to that described for the high school clusters. Desegregated and segregated schools are defined as before. Schools are stratified into appropriate SES strata, clusters are identified, and randomization is employed to select clusters for the sample.

11. ANALYSIS OF RESEGREGATION DUE TO OUT-MIGRATION FROM CENTRAL CITIES

Perhaps the single question asked most often by local policy-makers about school desegregation is "How many Anglo parents move from the city, or transfer their children to private schools as a result of desegregation?" The assumption is frequently made that white flight is a serious problem; in fact, majority families (and sometimes minority families as well) have been moving to the suburbs in segregated metropolitan areas, too, and we do not know whether there is greater movement in desegregating cities. Local policy-makers have a legitimate interest in this question. If most of the majority students disappear from the school system, desegregation will lead only to resegregation. If higher-income taxpayers move out of the city or send their children to private schools, there is a danger that financial support for the public schools will decay.

This study is designed to answer two questions: "How large is the withdrawal from the public school by either Anglo or middle-class minority parents as a result of desegregation?" and "If there are any cases where this withdrawal has been large, what sorts or policies can be used to reduce flight from the public schools?"

We propose a two-stage analysis. First we propose a careful analysis of a number of factors previously neglected in most research in this area, but excluding school desegregation policy and strategy. Second, we propose to add to the first analysis a set of variables measuring school desegregation policies and strategies in order to determine whether or not they have an effect above and beyond the effect of demographic, economic, and ecological factors. In the development of the second phase of the analysis—the policy-related phase—we also suggest adding an analysis of white flight via public to private school transfers, black flight via within-city or city-to-suburb movement and transfers to private schools, and attempting to relate these general and long-term processes to racial integration and turnover within neighborhoods. We believe, although there is little quantitative empirical evidence, that desegregation of schools can operate favorably for both the integration and racial stabilization of neighborhoods.

This Add-on Model is based on the expectation that whatever the impact of school desegregation on racial changes in central cities, it will necessarily be small compared to other forces. This is shown dramatically in Fig. 11.1, where we compare the rate of decrease in the white population of the public school population of Washington, D.C., before 1954, when the schools were *de jure* segregated, to the years immediately after, when Washington's schools were repeatedly singled out as examples of the failure of desegregation. Washington schools became predominantly black four years before the schools were desegregated, and a simple continuation of the overall trend of the pre-desegregation era would be sufficient to produce Washington's present nearly all-black school system. The data suggest that desegregation had a small effect in increasing white flight; but even this conclusion cannot be drawn until we rule out the possible impact of changes in the economy and the Washington housing market between 1949 and 1958.

This part of the research program would utilize existing published data from the U.S. Census and the Office of Civil Rights of the Department of Health, Education, and Welfare, along with data collected from the parent's survey described in Sec. 6 above, and would provide information for selecting sites for case studies.

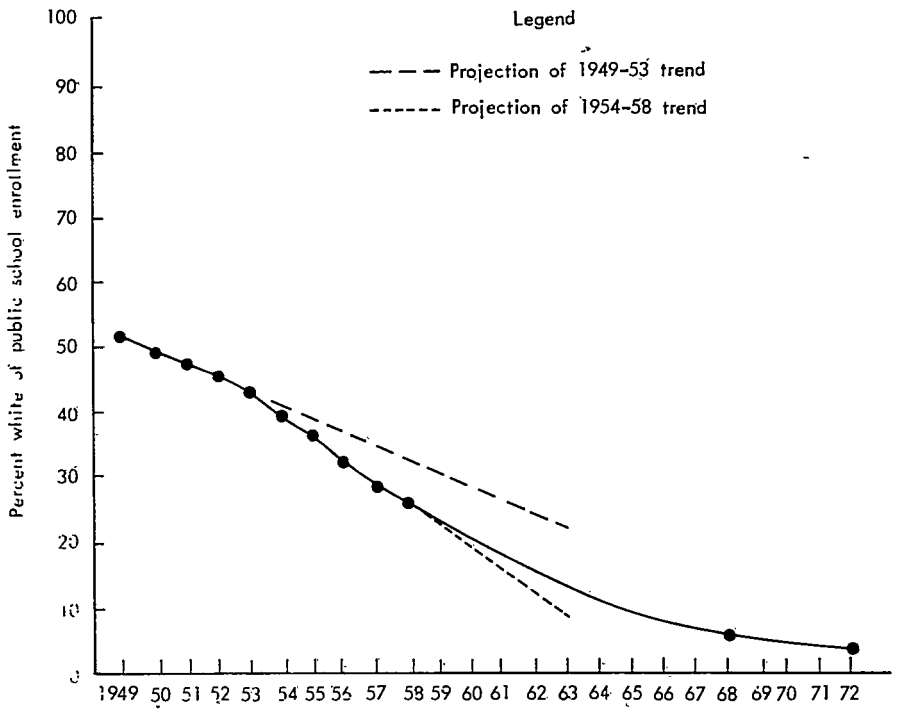


Fig. 11.1—Racial composition of Washington, D.C., public schools, 1949 to 1972, with projections based on pre- and post-desegregation trends

We believe that such an analysis would provide policy-makers with information nowhere else available on the viability of different desegregation plans.

The demographic history of the post-World War II period has been one of metropolitan growth, along with black in-migration to the central cities of the SMSAs and white out-migration to the suburbs of these SMSAs. The fact that U.S. cities have become more residentially segregated during this period is well-documented (Taeuber and Taeuber, 1965). There is also consensus on *some* characteristics that predict differing rates and types of suburbanization (Schnore 1973; Downs 1970). However, little agreement exists and little analysis has been done on the relationship between central city segregation and suburban migration. Literature and scientific consensus are scarcer on the question of how central city racial composition of schools relates to suburban migration and/or housing integration. Although two recent articles (Farley 1973; Farley and Taeuber 1974) discuss the issue, neither attempts a comprehensive analysis nor develops a well-defined analytic model.

Thus, a first phase of the analysis we propose must be the development of an explanatory model of factors that explain differences between one SMSA and another in racial patterns of city-suburban residential settlement.

THE ANALYTIC MODEL

The first problem in piecing together an analytic model and a research design is defining the dependent variable. In the simplest terms possible, the dependent

variable is the intermetropolitan variation in rates of out-migration from central cities to suburbs. It is measured over the period 1960 to 1970 because this was the decade of most desegregation activity and because we have to rely on existing data for that time period.[1]

The first qualification we must put on the above definition of the dependent variable is that we are interested in families with children of school age or pre-school age. We are not interested in single individuals or families with grown children. This qualification leads us to a second somewhat more complicated qualification. We know that city-to-suburban migration is more likely to take place among families with young children than among other households. We also know that this is largely due to the perceived quality of schools and housing available in the suburbs. Families with young children clearly have a preference for single-family units, with multiple bedrooms, and a lawn or yard.

The problem becomes a question of whether the housing supply—both existing stock and new housing—in the city and the suburbs is at all similar. One can take two approaches to this problem. One can argue that characteristics of the housing supply are determinants of within-SMSA migration patterns and thus should be incorporated into the analytic model as independent variables. A second approach is to begin with the expectation that SMSAs differ on whether or not *any* similar housing possibilities exist in both city and suburbs. This argument is that some SMSAs have dual housing markets in the city and suburbs with households participating in only one or the other, while other SMSAs have a relatively unitary housing market—or at least housing markets that interact and households that move from one to another. The second approach implies that it is invalid to compare SMSAs that have single or interacting housing markets with SMSAs that have dual or separate housing markets. Thus this approach would suggest two separate analyses of these two different types of SMSAs. The second approach is analogous to the dual labor market approach used in much recent analysis of urban poverty and developing economies. It derives from simplistic comparison of older, usually eastern cities and SMSAs with newer western and southern cities and SMSAs.

The second approach to the definition of the dependent variable above pertains solely to the distinction between the city and suburban housing supply. An additional qualification—also similar to and suggested by much analysis of not only dual labor markets but also dual housing markets—pertains to the degree to which cities and suburbs are open or not open to blacks and other minorities. In its simplest terms, this qualification may be whether or not any minorities live in suburban areas, although that distinction is probably too rigid. More realistically, it is a distinction between SMSAs in which there are integrated suburban communities, and SMSAs in which there are not. Both criteria should be used. Both are important because they symbolize the possibilities open to both blacks and whites, because availability of suburban housing to blacks lends a different significance to central city school desegregation, and because pioneering is much more difficult than continuing a process such as suburban desegregation. It seems that SMSAs where minimum breakthroughs in suburban housing desegregation have taken place are quite different from SMSAs where no breakthroughs have occurred.

We expect that the two characteristics of suburban vs. city housing supply and minimum desegregation are of such importance in determining out-migration patterns of whites that analysis should be carried out separately in different kinds of SMSAs. However, the evidence does not presently exist to establish this importance,

[1] Ideally we would want to examine more recent information and would want to include consideration of what happens during the time period of the panel study. Some less global analyses such as these will be suggested later in this section.

so we are recommending that both approaches be used: (1) an analysis of white family out-migration, including these variables in regression analysis and (2) an analysis that looks at characteristics affecting white out-migration separately in these different types of SMSAs. Below, we first outline very briefly the procedure for separating SMSAs in the latter approach and then describe important variables to be included in both analyses.

What was described above amounts to a fourfold classification of metropolitan areas according to differences between the central city and the suburbs within it. This classification is described in Fig. 11.2.

	Housing Supply: Single or Somewhat Similar	Housing Supply: Dual or Very Dissimilar
Some blacks in suburbs and racially mixed communities in suburbs	(A)	(B)
No blacks in suburbs or no racially mixed communities	(C)	(D)

Fig. 11.2—Racial composition of city vs. racial composition of suburbs

To classify SMSAs into one of the four cells in Fig. 11.2, we would first do a cluster analysis of housing characteristics, including size, cost, single dwelling vs. multiple, and density, both in terms of population per square mile and people per room. We expect that these would cluster for small geographic areas—tracts within central cities and suburban townships. To the extent that this clustering indicates distinct neighborhoods with regard to housing supply, we can then compare central cities and suburban rings with regard to how much they differ on these clusters of characteristics. We expect that the differences will be clear enough so that we can then classify SMSAs in terms of those in which city housing supply is different from suburban housing supply and those where it is not different. Census data can be examined for information on whether or not blacks live in the suburbs. We can examine census materials again for the question of whether smaller units—such as towns, groups of tracts, or even tracts—are racially mixed. When these are found, we can use local informants and researchers to find out whether this is an artifact of boundary lines of the units or represents some residential integration. With this information, SMSAs can be classified according to the vertical dimension above.

How do we expect this classification scheme to be related to white out-migration to the suburbs? We expect the general rate of out-migration to rank from highest to lowest in the following manner:

Highest out-migration rate (1960-1970)	D Cell
Second highest	B Cell
Third highest	C Cell
Lowest out-migration rate	A Cell

However, we expect the *association* of other factors and especially level of school desegregation with white out-migration to follow a different pattern:

Highest association between level of school desegregation and white out-migration	C Cell
Second highest	D Cell
Third highest	A Cell
Lowest association between level of school desegregation and white out-migration	B Cell

THE ANALYSIS PLAN

The analysis strategy is a regression-type approach with which it is hoped to build a path model of factors affecting white out-migration and locate school desegregation within that path model.

The first phase of such a strategy involves analysis of variables (other than school desegregation) that would affect the rates of white residential out-migration. The dependent variable would be based on 1960 and 1970 racial composition of the total population. These variables, listed below, would also be used in analyses conducted within the categories of the typology above; variables that form the typology are indicated in parentheses.

1960 to 1970 White Out-Migration = Function of:

(Similarity of housing supply in city vs. suburbs)
(Number of blacks and other minorities living in suburbs)
(Number of blacks and other minorities living in mixed communities in suburbs)
1950-1960 and 1960-1970: Growth rate of SMSA jobs, population, other economic indicators
Rate of black and Spanish-speaking minority in-migration, 1950-1960
Rate of black and Spanish-speaking minority in-migration, 1960-1970
Residential segregation index (Taeuber and Taeuber 1960)
Change in residential segregation index: 1950-1960, central city (Taeuber and Taeuber 1960)
Residential segregation index, 1970
Change in residential segregation index, 1960-1970
School desegregation index, 1967
Density within minority census tracts: overcrowding of dwelling units
Density within minority census tracts: population per square mile
Density within white census tracts: overcrowding of dwelling units
Density within white census tracts: population per square mile
Proportion of total SMSA population within central city
Proportion of total SMSA jobs within central city
Existence of stable (1960-1970) mixed census tracts within city
Age of housing in central city

Difference in age of housing, central city vs. suburbs
Regions of country
Rate of change in black and other minority students in public schools, 1967-1970
Indexes (Gini) of income and other SES inequalities between minority and majority populations
Proportion of blocks changing from white to black (or Spanish-speaking minority) occupancy, 1950-1960
Proportion of blocks changing from white to black (or Spanish-speaking minority) occupancy, 1960-1970
Presence of desegregated suburban school districts
Crime rate of central city and suburbs (uniform crime reports)

Additional variables may be added to this list by the research team. Since this analysis uses only census data, it can use either the screener sample or, if warranted, the universe of 1399 school districts from which the screener sample is selected. The bulk of the analysis would be concentrated on the central cities of the sample.

Having estimated the effects of these factors on white out-migration, it would then be necessary to estimate the additional effect—positive or negative—of different levels and strategies of school desegregation. These would include the following:

Proportion of minority (by race) students desegregated
Proportion of white students desegregated
Whether school desegregation is court-ordered or voluntary
Whether school desegregation involves busing and how many students of each race are bused
Whether grades and schools are reorganized
Whether public officials support school desegregation
Whether the media support desegregation
Whether the school system attempts a public information effort
Whether teacher and staff preparatory training is undertaken
Number of whites sent to schools previously identified as minority schools.

Part of this analysis would be done using the screener sample; other parts would use only the district leadership survey sample.

Two additional analyses would be done next.

First, the above analysis should be redone, using estimates of the change in racial composition of the public school enrollment. This introduces the important consideration that white flight or minority flight might occur not only through residential migration out of the central city but also through transferring children from public to private schools. Student racial composition is not available for 1960. The proportion of students in the public schools who are white can be estimated as a function of proportion of families in the city who are white and the proportion of students enrolled in private schools for 1970. Using this 1960 estimate and the 1970 actual figure, one could compute change in the proportion of nonwhite students as a residual of the change in proportion of the nonwhite population from 1960 to 1970. Having done this, one can then use the same variables described above to analyze

characteristics of SMSAs and of the desegregation process as they relate to this estimate of white withdrawal from public schools. This will give an independent, though far from satisfactory, estimate of private school withdrawal. Using the typology developed for characterizing SMSAs in terms of housing supply and city-suburban racial differences, one could determine if in some SMSAs housing migration is constrained and private schools are used as an alternative to escape desegregation.

The second additional analysis would use information from the parent surveys included as part of the community reaction study. These surveys will ask some questions concerning probabilities of residential relocation and school racial composition. However, probably more important than those questions is the use of those surveys to make city level estimates of perceptions and attitudes of the residents. For example, we included the central city crime rates as a predictor of suburban flight above. However, we also know that, in addition to the uniform crime reports being subject to error, people's perception of crime as a problem is not simply a function of the crime rate—however accurately it is measured. In fact, perceptions are probably more important than reality for predicting behavior with regard to suburban migration. We hope to add the following list of estimates to the regression analysis described above in the 100 northern cities and 70 southern cities in which we do parent surveys:

Perception of the extent to which crime is a problem
Perception of "quality of life" in city
Perception of quality of schools in city/neighborhood
Evaluation of school board and school administration

Since parental data are highly clustered within each city, their predictive power can be considerably improved if they are first standardized by pooling the entire sample, estimating the effects of personal characteristics such as age, SES, and number of children, and then adjusting attitudes for each city as if the means on these predictor variables were equal the cities' overall means projected from the 1970 census.

The theoretical model is briefly summarized in Fig. 11.3. We have not recommended an exact statistical procedure for the analysis, since it is not expensive to

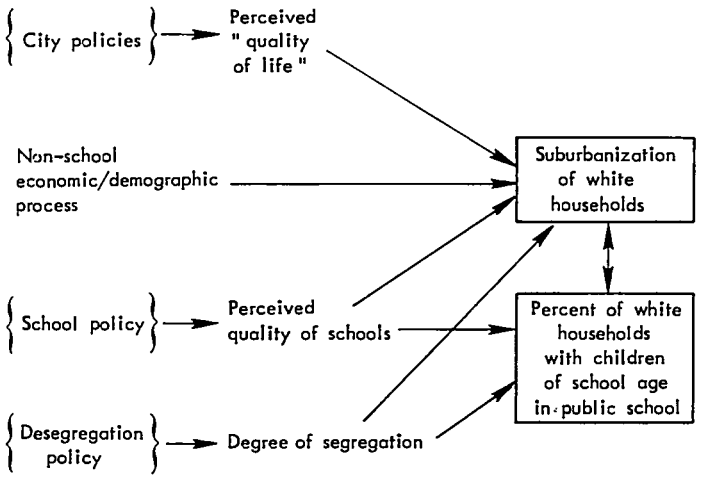


Fig. 11.3—The theoretical model

perform the analysis in a variety of ways. In general, we favor using regression to estimate the effects of demographic and economic factors in white flight in a large sample of districts that have not desegregated; data for desegregated districts can then be analyzed fixing these effects and permitting desegregation plan characteristics to affect only the remaining unexplained variance. Proceeding in this fashion permits very large sample sizes to be used for the first part of the analysis, which does not use parent or community interview data.

Table 11.1 suggests one possible analysis plan, which first analyzes the migration of households with children in a 5-step process, and then repeats the analysis, using ethnic composition of the schools as a second dependent variable. At each step in the process, the maximum amount of information from the preceding steps is used.

This analysis should be performed separately for communities where the minority group is black and where a substantial Spanish-speaking minority exists; a very similar analysis plan can be used to analyze the out-migration of middle-class minorities from central cities.

Table 11.1

SAMPLE ANALYSIS PLAN

Equation No.	Sample	Dependent Variables	Independent Variables
1	Either screener or all cities with 3000 minority population in SMSAs	Change in ethnic composition of households with school age children, 1960-70	Demographic and economic variables
2	Screener sample	Change in ethnic composition of households with children, 1960-70	Add on pre-1970 desegregation plan data to Eq. 1
3	District survey	Change in ethnic composition of households with school age children, 1960-70	Add on desegregation, community factors to Eq. 1
4	Districts with sampled schools	Change in ethnic composition of households with children, 1960-70	Add on parental nonschool attitudes to Eq. 1
5	Districts with sampled schools	Change in ethnic composition of households with school age children, 1960-70	Add on parental school and nonschool attitudes to Eq. 3
6	Either screener or all cities with 3000 minority population in SMSAs	Change in ethnic composition of public school students, 1968-present	Demographic, economic, private school attendance 1960, and household migration, 1960-1970
7	Screener sample	Change in ethnic composition of public school students, 1968-present	Add on desegregation plan data, 1968-present to Eq. 1
8	District survey	Change in ethnic composition of public school students, 1968-present	Add on desegregation plan community factors to Eq. 1
9	Districts with sampled elementary schools	Change in ethnic composition of public school students, 1968-present	Add on parental nonschool attitudes to Eq. 1
10	Districts with sampled elementary schools	Change in ethnic composition of public school students, 1968-present	Add on parental school and nonschool attitudes to Eq. 3

12. EXPERIMENTAL DESIGN IN SCHOOL PROCESS STUDIES: A PREFERRED ALTERNATIVE

At various points in this report, we have noted that experimental design using the random assignment of treatments to schools is considerably superior to the simple longitudinal designs we have proposed. However, we have not recommended experimental randomization because of considerable sentiment among our advisers that the nation is not yet ready to adopt such a new approach to evaluation in education. As we are not convinced that this is true, in this section we discuss why an experimental design is a preferable approach. We then suggest two ways the research program can be modified: (1) by converting the process substudies of Sec. 9 into an experimental design and (2) by converting the large-scale study (Sec. 7) into an experiment.

WHY ARE EXPERIMENTS WITH RANDOMIZATION A PREFERRED METHOD?

The scientific community recognizes that randomized experiments, wherein a "treatment" is randomly assigned to some schools in a sample while the others remain as "controls," is the only absolutely correct method for evaluating the effectiveness of a treatment. While this method is routinely used in the physical sciences and medicine, its use in social science and social policy research is newer, and only one major experiment has been executed in the federal evaluation of educational programs.

Experiments with random assignment permit isolation of the effect of treatment from the effects of all other variables, such as student socioeconomic status or associated school factors. For example, if we attempted to evaluate the effects of remedial reading efforts without a randomized experiment, we would need to somehow make allowance for the fact that students who receive remedial assistance are likely to be of lower academic achievement, and that schools with well-developed remedial reading programs will be different than schools without them in a variety of ways (perhaps having more dedicated teachers, higher levels of Title I funding, etc.). The first problem can be anticipated and various statistical techniques used to produce an artificial matching of achievement levels of students in remedial reading programs with those of students not in such programs. Such methods have serious technical difficulties and have been widely criticized; this criticism is central to the debate over the validity of the Coleman Report analysis. The second problem—"control" on all other relevant school factors—is an insoluble one unless we are willing to assume that some researcher exists who can imagine every significant way in which the one group of schools differs from the other and can measure all of these differences accurately in order to control for their effects. A third problem, less often noted in the literature, is that without the experimental assignment of treatments to schools we must determine whether the school has a particular treatment or not; anyone who has studied schools knows that it is a difficult task to determine the size of a remedial reading program, a description of it, and sometimes even whether it actually exists.

These methodological difficulties are serious in practice as well as in principle. Perhaps the strongest evidence of this is the meager number of conclusions that

have resulted from the vast amount of educational research over the past decade. Granted, many educational research projects during that time were small, and some poorly executed, but even *Equality of Educational Opportunity,* a well-financed study done with reasonable skill, resulted in more controversy than definite conclusions and to this date some of its major findings are not generally accepted. Many research studies have been carried out on school desegregation, but methodological problems are so serious that the overall result has been disagreement rather than agreement about the effects of school desegregation. Many studies of the effects of school desegregation have been longitudinal, which does represent a considerable improvement in methodology over one-time cross-sectional studies such as the Coleman Report; however, the mere existence of a longitudinal study does not mean that problems of spuriousness or bias are solved, as indicated by the intense controversy over the effects of Headstart (Campbell 1971).

In contrast, the first experimental evaluation, the National Opinion Research Center evaluation of the Emergency School Assistance Program, used a randomized assignment of ESAP funds to schools and concluded that black male tenth graders had higher achievement test scores as a result of the program. At the same time the NORC team used conventional regression methods to analyze a large number of school characteristics and were able to locate only one school innovation (the use of media) which seemed to positively affect achievement to the extent that ESAP did. It seems unlikely that the presence of ESAP funds, which averaged $10,000 per school for one year, represented the largest difference between these schools. We believe that other school factors should be as important as the presence or absence of ESAP funds, and standard regression techniques were too weak to locate them.

To pursue one example from the NORC survey further, that study attempted to evaluate the effects of remedial reading programs, by looking at achievement test scores in schools where principals said a remedial reading program was being used and also in schools which had a remedial specialist on the staff. First, the effects of having a "remedial reading program" were completely different from those of having a "remedial reading specialist," pointing up the difficulties of determining whether a school has a particular treatment or not. Second, the simple correlation between the presence of a remedial reading specialist and the achievement test scores of black students was −.19, indicating, as one might hope, that remedial reading specialists were assigned to needy schools. When an effort was made to control out the effects of differences of social background of students, this correlation increased considerably, although not enough to become positive; the final standardized regression coefficient was approximately zero. In reviewing the research, the study director and the project monitor point out that when they attempted to correct statistically for the measurement errors in the study, they estimated that schools with remedial reading specialists may be able to increase achievement by an average of one-half grade level, with measurement error reducing this positive effect to zero (Crain and York, forthcoming).

In summary, conventional educational research to date has been unable to match students in treated and untreated schools, and unable to measure all possible additional factors that might be a true explanation for an apparent effect or noneffect of a treatment. All efforts to statistically match students are subject to serious methodological difficulties, and in the study cited above, the authors recognize the possibility that these errors had made a true achievement gain disappear.

In contrast, it is a technically simple task to randomly assign treatments to schools, knowing that, thanks to randomization, the schools can differ on no other characteristics except the presence of the treatment, within the limits of sampling error. We conclude that in practical application the use of randomized experiments

represents a considerable gain in the possibility of drawing policy-relevant conclusions.

The major difficulty with experimental treatments is a social or political one. In our case, it means that school systems must agree that a flip of the coin will prevent certain schools from receiving a particular program. One can well imagine that had someone proposed randomly allocated Title I funds, school staff and parents alike would complain that randomization was tantamount to random discrimination. Charges of discrimination or favoritism are the most frightening of all to federal program administrators. In the case of the ESAP evaluation, the charge of discrimination could not be made, since the underfunded program would have had to omit most schools, whether by administrative decision or by randomization. Nevertheless, a number of school superintendents refused to cooperate with the experiment. On the other hand, it should be noted that among those 103 school districts that did cooperate, there was virtually no indication of any community protest about the random assignment of federal funds.

In the remainder of this section we describe a method of allocating experimental treatments to schools which we believe will succeed in reducing political and community objections. The goal of the experiment program is to develop, field-test, and evaluate several innovations in such a way as to bring the innovations close to the stage of mass dissemination.

We are not concerned with developing genuinely radical school innovations. Rather, we propose a research program to consolidate the developments already under way, and to pave the way for mass acceptance of those innovations which are now being used successfully on a small scale.

In the following subsection we discuss the application of experimental methods to the substudy of desegregated school processes of Sec. 9. We later propose a method for adapting the large-scale longitudinal study of Sec. 7 to experimental methodology.

THE EXPERIMENTAL VERSION OF THE SCHOOL PROCESS SUBSTUDY

There are three basic questions to be answered:

- How should treatments be selected?
- How can the cooperation of schools be obtained?
- How shall the innovations be implemented?

We believe that educational literature and present educational practice contain a number of successful innovations. Unfortunately, present educational practice no doubt also contains an even larger number of unsuccessful innovations, so that we cannot recommend the mass adoption of particular programs now. We can, however, select from the existing practices in schools a number of promising treatments. We recommend that these treatments be designed, not by theoreticians or university researchers, but by school administrators—ideally, administrators and principals who have demonstrated their ability to operate unusually effective schools. Our first recommendation then is that a design competition be held among administrators of unusually effective schools in which they are invited to submit outlines of the innovative practice that they would recommend for adoption by the schools. Unusually effective schools would be identified from student outcome data in the large-scale longitudinal study. Contracts would then be awarded to the winners of the design competition to enable them to develop completely detailed designs.

In order to minimize problems of bias and to ensure that the proper mix of technical skills is present, five other groups need to participate with the innovation designers in the experiment.

- The research team analyzing the large-scale longitudinal study selects unusually effective schools prior to the design competition, uses their analysis to help select innovations, and draws the sample of schools for the experiment.
- A technical assistance team implements the innovations by placing staff persons in each school.
- The principals and district administrators of the participating schools work together (on a paid basis) in a congress which provides technical assistance to both the designers and the technical assistance team.
- The evaluation should be carried out by a separate research team; since they are unlikely to have experience in data collection, a separate data-collection team (perhaps the team that collected the data from schools in other parts of the research program) should be used.

We believe that dividing the work among these six different actors will minimize the two most serious problems: the control of experimenter effects and the willingness of schools to cooperate.

It is important that we evaluate the design itself, and not the charisma of the designer. Some school administrators may be superb personal leaders who could develop a given program into a highly successful one, but the same program might not be "replicable"; in the hands of another administrator it might fail badly. It is therefore important that the treatment be implemented through the use of an impartial team of technical assistants, with the actual technical assistance staff assigned to schools on a randomized basis, thus eliminating the possibility of bias in the execution of the treatment and permitting us to evaluate the treatment exclusive of variation in administrative techniques or skill.

We believe that school administrators will not be reluctant to participate in the experiment, partly because they can have hope of improvement in schools receiving the experimental treatment, and partly because they would receive additional funds in connection with the experiment. Finally, we recommend that school administrators be involved in the planning of the project as paid consultants from the beginning of the project, and that they be brought together in a series of group meetings to discuss their feelings about the experiment and to recommend improvements.

Figure 12.1 indicates how the six actors are involved in the seven steps of the project.

Selection of Schools

The design assumes that 20 schools should be tested with each innovation, and that six innovations will be developed and considered worthy of trial. Since some of these innovations will be restricted to certain grade levels while others will be generally applicable to both elementary and secondary schools, we conclude that approximately 80 schools should be experimentally treated at each grade level. Since we can anticipate that perhaps 20 schools may withdraw after the experiment has begun, and since we will need a control group of approximately 25 schools at each grade level, our final sample size should be approximately 375 schools—125 high schools, 125 junior high schools, and 125 elementary schools.

Given our interest in developing innovations suitable for newly desegregated schools (and given our additional interest in observing the process of change in desegregated schools, which we will be able to do within the context of the experi-

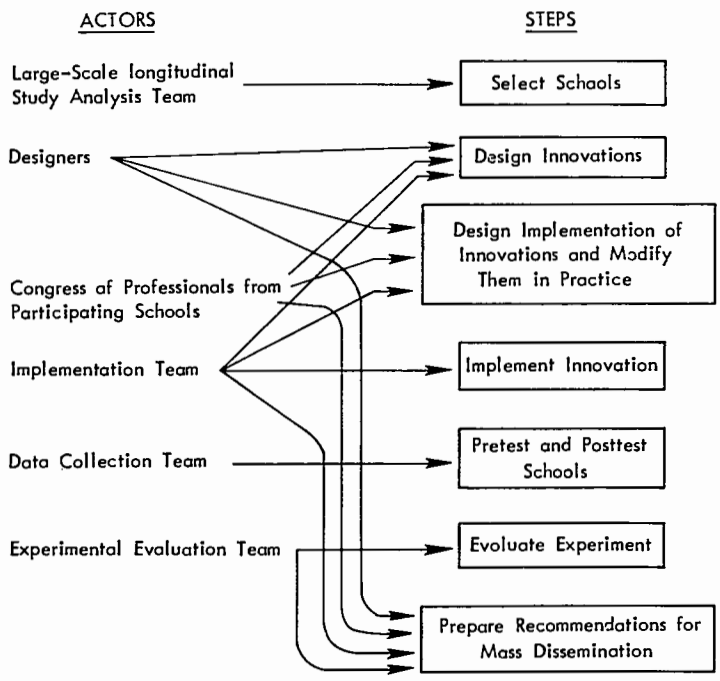


Fig. 12.1—The experimental design

ment), we should oversample schools in the early stages of desegregation. These schools should be assigned equally to each treatment, so as to eliminate this source of bias from the evaluation.

The more we minimize differences in student outcomes at the time of the pretest, the more powerful our evaluation will be. Therefore the analysis of the longitudinal study should include the identification of matched sets of schools, each set containing one more school than the number of treatments planned. Thus, if four treatments are planned, 25 sets of five each should be selected, matched on the overall level of achievement, quality of race relations, racial composition, etc., and also matched on the rate of growth of the student outcome variables during the one year of the large-scale longitudinal study. With losses due to attrition, this sample would suffice. In order to remove the simple effect of financial aid, the control schools will be awarded funds equal to the average spent on the innovations with no strings attached.

Selection of Innovation Designers in a Design Competition

Innovations should be designed by experienced people, familiar with the problems of schools, experienced in the administration of schools, and with proven capability to produce educational results. The large-scale longitudinal study, conducted during the first two years of this research program, provides a ready source of such persons; namely, the administrators of unusually effective desegregated schools. Therefore, one of the first steps in the analysis of the large-scale study should be the location of unusually effective schools. These principals, and their supervising administrators, should be invited to participate in a design competition. The designs submitted should be evaluated by government officials and by desegre-

gation experts from universities, who would compare their proposals with the results of similar projects in the large-scale longitudinal studies. We anticipate that in nearly every case it will be possible to perform a "quick and dirty" evaluation of these proposed innovations. If a school principal participates in the design competition (whether his or her design is accepted or not) the participants' school is excluded from the experiment.

Assistance to the successful candidates in this competition should be provided from two groups to be discussed below: a congress of professionals from the schools to participate in the experiment, and the implementation team, a group selected to actually execute the innovations in the sample schools.

Design of Implementation of the Innovation

The congress of professionals from participating schools should be selected and invited to a first meeting as soon as the sample is selected, even before cooperation from their school districts has been obtained. After they have been informed of the overall projects, they can begin participating in recommending techniques for implementation. The designers of the innovations can work with subcommittees from the congress in developing their designs, and subcommittees of the congress can work with the implementation team in developing the plans for implementation.

The implementation team should be a group experienced in the development and execution of school innovations. The team should employ a staff of 80 "technical assistants" who would spend most of their time providing technical assistance to the schools as they implement the innovation. These technical assistants should be experienced teachers (or school administrators). Their function is similar to that of a community organizer, and they should have similar qualifications. In order to minimize the effects of personal characteristics, we recommend that each of these implementation team technical assistance specialists be assigned to supervise three different innovations in three different schools and that the innovations be assigned randomly. In order to assess the import of the assistance team, one-fifth of the schools should not receive this assistance. This would minimize the effects of personality on the success of the innovations. Depending on what innovations are used, additional staff may or may not be necessary in the schools. Of course, existing staff should be used as much as possible. We therefore recommend that the summer before the experiment starts be used as a paid training session for selected teachers to enable them to execute their portion of the innovation.

Implementation of the Innovation

We recommend that the innovations operate for three years. Minimal modifications in the program can be made on the basis of first-year results, but no public report should be made at that time. An interim report can be made at the end of the second school year, when longitudinal results for two successive cohorts are available to compare with longitudinal results obtained in the large-scale study before the innovation was introduced.

It should be noted that the congress of professionals from participating schools will participate in design of the innovations and in design of the implementation without knowing which innovation will be executed in their schools. The actual selection of schools by random assignment to the treatments will not occur until the

very end of the preparation phase, only as early as is necessary to permit training of school staff.[1]

School Testing and Evaluation

The team to evaluate the experiment should be separate from the design and implementation groups. Experimental evaluation would consist of pre-tests and post-tests of succeeding cohorts at the same grade level, the pre-test occurring the spring before entering the grade, the post-test at the end of the school year the following spring. Analysis can be a traditional analysis of covariance, and should involve no special difficulties. Where schools withdraw from the experiment due to local objection, pre-testing and post-testing should continue so as to avoid introducing an additional bias into the experimental analysis. Thus if a particular treatment is executed by 25 schools initially and five withdraw, the evaluation treatment should still be based on all 25 schools, the losses being considered only after the statistical analysis is complete. Any alternative approach introduces bias which covariance cannot totally correct for.[2] The additional special problem in evaluation is the presence of Hawthorne effects, and what we might call learning effects.

The implementation team plays a role in the evaluation by preparing descriptions of how the designs were actually implemented. In addition, members of the evaluation team should visit the schools to observe both how the innovation has been accepted and how it was modified in practice, and to observe the effects of the innovation on the school's social climate.

The evaluation should look for treatment-school interactions; it may be that certain innovations are practical for certain types of schools and not for others. The use of the matched-set design should make analysis of interactions possible.

Preparation of Recommendations for Mass Dissemination of Favorable Results

At the end of the third year, the experimental evaluation team will report back the relative performance of schools in each treatment group. At this point, designers of successful treatments, members of the congress of participating schools, members of the implementation team, and members of the evaluation staff can work together to prepare a package to be made available to other schools interested in adopting the innovation. The overall schedule shown in Table 12.1 indicates that final dissemination of the innovations will occur in time for adoption during the eighth year from the beginning of the study.

ADAPTION OF THE LARGE-SCALE LONGITUDINAL STUDY OF SCHOOLS TO THE EXPERIMENTAL METHODOLOGY

The use of the experimental design outlined above in the process studies during the third through fifth years of the research project instead of during the first two

[1] The random allocation of schools will be a painful experience for the congress as most of them discover that their favorite innovation will not be used in their school. We recommend that a minimum of arcane computer technology be used for the randomization. The mass-selection of colored balls from Bingo cages would be preferred. If "n" innovations are used, schools are selected in geographically clustered groups of size n, and it would be feasible to form working teams of size n throughout the duration of the congress of representatives. Thus the randomization would take place among the "n" members of that team, each obtaining a different program.

[2] See the NORC ESAP report (NORC, 1973) for a discussion of the problems generated by attrition of this kind.

Table 12.1

SCHEDULE FOR EXPERIMENTAL STUDY OF SCHOOL INNOVATIONS

Year		Item
1	W	Notify schools of design competition when first contacted
	Sp	Pretest large-scale survey
	Su	
2	F	
	W	
	Sp	Posttest large-scale survey
	Su	Select participants for design competition
3	F	
	W	Select experimental schools; select innovations
	Sp	Pretest experimental schools, complete implementation designs
	Su	Begin in-service education in participating schools
4	F	Begin experiment
	W	
	Sp	First cohort posttest, second cohort pretest
	Su	Modify experimental designs
5	F	
	W	
	Sp	Second cohort posttest, third cohort pretest
	Su	Modify designs, terminate poor designs
6	F	
	W	Submit preliminary report of two-year experiment
	Sp	Third and final posttest
	Su	Begin final report, due spring of year 7; begin plans for mass adoption, fall of year 8

years of the project in the large-scale longitudinal study is preferable for only two reasons. First, the use of the large-scale longitudinal study as a pre-test permits a preliminary analysis of feasible innovations. Second, analysis of the large-scale longitudinal study permits us to select school administrators for the design competition. A third reason is much less important: analysis of the large-scale longitudinal study permits us to match the schools before randomization. However, randomization without matching is almost as effective, and the absence of pre-test scores can be compensated for by increasing the sample size.

How serious, then, are the two reasons for waiting for the third year of the study to design the experimental innovation? We think that a reasonably good group of candidates for the design competition could be selected by nominations of highly regarded school administrators. Each superintendent in the sample might be invited to submit the name of the candidate, knowing that a preliminary screening would eliminate a number of these before the design competition began. It is true, however, that any efforts to select successful innovations purely on the basis of their reputations would be quite inefficient. It is likely that the errors would be errors of commis-

sion rather than omission; in a large design competition, almost every successful innovation is likely to be proposed, the problem being that a number of unsuccessful ones might be selected as well as the successful ones.

This suggests that the following strategy might make it possible to adapt the large-scale longitudinal study to an experimental methodology. Rather than selecting only a half-dozen innovations in the competition, a larger number—perhaps 12—should be selected, and the implementation should take place in a larger number of schools. It is not necessary that every school in the large-scale study be included in a treatment, although for political reasons it would probably be advantageous to provide either a treatment or compensating funds in control schools for every school in the sample. To do this would involve an increase in cost during the first two years of the study, and would also require delaying initiation of data collection for one year while the innovations are being prepared and the schools assigned randomly.

Costs of the treatments to be executed will be in excess of $10 million. Nearly all of these costs are in direct aid to schools, and consequently a fair cost accounting would attribute these costs to the general federal aid program. Viewed in this way, $10 million represents a very small fraction of the annual budget of the Emergency School Assistance Act, for example.

Execution of the experimental version of the large-scale longitudinal study would follow almost exactly the plan described for the experimental process studies, beginning with sample selection, proceeding to a design competition, a congress of participating professionals, an implementation team, and finally the evaluation itself. Since the results would be for only one year, it is recommended that a number of the more promising innovations be continued for a second year. The school process studies would still be necessary. The one-year longitudinal study provides an inadequate test of the effectiveness of an innovation, since in many cases the start-up period is a poor time for evaluation.

Thus the experimental version of the large-scale longitudinal study does not affect the remainder of the design: the third year would still be used to develop more refined versions of the original innovations. However, the refined versions of the innovations would be greatly strengthened by the existence of the experimental data, and there is a moderately good chance that evidence from the first-year study will be strong enough to merit the recommendation to other school districts that they adopt certain types of programs. Recommendations for national policy were made on the basis of the ESAP evaluation, which was only one year long; the study proposed here would be superior to that one.

An additional argument in favor of the experimental treatment is that providing funds to school districts encourages the districts to participate in evaluation. In addition, the opportunities for school officials to participate in the design competition and to serve on the congress of professionals are additional incentives for school districts to participate.

Implications of the Experiment for Other Types of Analysis and Data Collection

If the experimental version of either the school process study or both the large-scale study and the school process study are adopted, we recommend that no changes be made in the remaining portions of data collection and analysis. Thus the experiment is in addition to, rather than instead of, the analysis and data collection outlined in Secs. 7 and 9. All analysis can be done as easily in schools experiencing an experimental treatment as in schools where no experiment takes place.

Use of the experimental design does permit some economies, since complicated subsampling can be simplified, and this will reduce the amount of data collection. This issue is discussed in Sec. 15.

Conclusion

The use of either of the two experimental designs proposed in this section will greatly increase the likelihood of locating viable innovations to improve desegregating schools. The design has two weaknesses. First, it is limited to conservative and easily accepted innovations; we do not think an experimental design is appropriate for more radical innovations, because it would be impractical to expect widespread acceptance by randomly selected schools. Second, the innovations may have somewhat different effects in the experiment than they would in mass use due to Hawthorne effects and experimenter effects.

The main problem with the experimental method is political. The experiment is a departure from past methods of evaluation and requires a commitment by federal agencies to a politically new idea. We believe that such a federal commitment would prove to be worthwhile.

13. EFFECTS OF SCHOOL DESEGREGATION ON POST-HIGH SCHOOL OUTCOMES

A good portion of this report has been devoted to the design of studies that investigate the effects of desegregation on students and staff within the school setting. Clearly, this is only the start of an evaluation of the effects of school desegregation; most policy-makers would probably place more stress on relatively long-term benefits that carry into and throughout adulthood. Accordingly, we place high priority on the investigation of desegregation effects on post-high school outcomes such as college attendance, career choice, and race relations. This priority gains more urgency when we realize that little research has been done in this field (Kapel 1969; Crain and Weisman 1972; Armor 1972), and what little there is does not settle the question of whether or not school desegregation leads to more beneficial career decisions by minority students.

Given these circumstances our original proposal stressed the importance of a longitudinal study that followed panels of desegregated and segregated students through several years of post-high school experience. Further thought and investigation has led us to conclude that a slightly different approach should be followed. While we still feel post-high school outcomes are crucial for a full evaluation of school desegregation, it is now apparent that there are a number of existing post-high school panel studies that might be fruitfully utilized to provide a preliminary look at selected outcomes. Post-high school panel studies are very expensive if they are to be done properly, and it seems to us that a new one should not be designed solely for the purpose of exploring the desegregation process until we have exhausted the available data. It is unlikely that these secondary analyses will answer all the questions raised by our models of the desegregation process, but their results should enable someone to design a much more refined study that would fill in the gaps. In the meantime, we feel that much valuable information will be produced for a relatively small cost.

In this section we will identify three sets of data and describe how they can be used to provide preliminary information on the effects of school desegregation on post-high school outcomes. These are not the only sources of data that could be so used, but they are the ones that offer the most promise.

PROJECT TALENT

Project Talent is a national longitudinal study of 400,000 students who were enrolled in grades 9 through 12 in 1960. These students are being followed up 11 years after they would normally have completed high school, i.e., between 1971 and 1974. Data will be available on their educational attainment, occupational status, earnings, and career plans. The initial sample probably included about 40,000 black students. By comparing black students who attended all-black high schools with those who attended high schools that also enrolled whites, it should be possible to estimate the impact of high school racial composition on subsequent life chances. The Project Talent data bank includes detailed information on family background, test scores, grades, curriculum assignment, and educational plans at the time of the initial survey, so it would be possible to determine if these factors are responsible

for any observed differences between blacks who attended the two sorts of high schools. The same comparisons can also be made between whites who attended all-white schools and those who attended racially mixed schools.

The data include race, sex, family background characteristics, various test scores, high school grade point average, high school curriculum assignment, educational plans at the time of the high school survey, college attended, highest grade completed, occupational status 11 years after high school, earnings 11 years after high school, career plans 11 years after high school, region, community size, high school racial composition, and a variety of other high school characteristics, such as teacher-pupil ratio and district expenditures per pupil. A fair amount of items concerning father's occupation, father's education, family income, and educational plans received no response.

It would be quite simple to regress various measures of adult success (highest grade of school or college completed, college quality, occupational status, earnings, career plans) on high school racial composition, controlling of course for such things as parental socioeconomic status. Correlations between student characteristics in high school (i.e., test scores, curriculum assignment, grades, and educational plans) and these measures of adult success could also be obtained, but it is not clear whether they are determinants of high school racial composition or determined by it.

This causal ambiguity could be reduced by separating ninth and tenth graders from eleventh and twelfth graders. If high school racial composition has an appreciable impact on a trait like test scores, the relationship should be stronger at higher grade levels. Conversely, one would hardly argue that high school racial composition could have significant effects on test scores in the ninth grade. If controlling ninth grade scores explained all observed differences between black adults who had attended segregated schools and those who attended desegregated schools, one would have to conclude that high school racial composition alone had little independent effect on adult characteristics. One could not, of course, rule out the possibility that elementary school racial composition had produced the test score differences between blacks in segregated and desegregated high schools.

One would want to design this regression analysis so as to capture nonlinear relationships between variables. This would be particularly important with respect to the effects of racial mix, since it is important to look for "tipping points." Where sample size permits, one would also want to stratify the sample in various ways to see if there were important interactions between the independent variables.

As mentioned previously, the Project Talent data bank is built around a stratified sample of students who were enrolled in grades 9 through 12 of American junior and senior high schools in 1960. The initial sample included 987 high schools, plus 238 junior high schools that normally sent their graduates to one of the 987 high schools. This sample was designed to be representative of all American secondary schools, public and private. Small high schools were undersampled and large high schools were oversampled. This bias can be corrected by appropriate weighting.

Ninety-three percent of the sample schools agreed to cooperate. The principal of each school filled out a questionnaire describing his school in some detail. Each student in the school was also supposed to fill out a lengthy questionnaire describing his family background, current activities, attitudes, and plans. Each student also took two days of tests covering a wide range of cognitive skills and information. These tests are identified, and questionnaires are included in the Project Talent handbook (American Institutes for Research 1972). A major problem in the analysis of these data is that students were not asked their race at the time of the initial survey, although principals were asked the racial composition of their school. There-

fore, inferences drawn from the initial survey data and the one-year follow-up to it must be based on estimates of representativeness.

According to the data-bank handbook, the Talent survey indicated that the racial composition of American high schools in 1960 was as shown in Table 13.1. (Note that the right-hand column is the estimated percentage of *schools* with a given enrollment, not the percentage of *pupils* in schools with that enrollment.)

Table 13.1

RACIAL COMPOSITION OF AMERICAN HIGH SCHOOLS IN 1960

Percent Black	Percent Schools	Percent Black	Percent Schools
None	68.2	50-59	--
0-9	20.4	60-69	0.1
10-19	2.4	70-79	0.1
20-29	0.5	80-89	0.1
30-39	0.2	90-99	0.3
40-49	0.1	All	7.4

These estimates are meant to describe the universe of American high schools, not the actual Talent sample. The Talent sample includes fewer small schools and more large schools than does the nation as a whole. In 1960, all-black schools were most likely to be found in the rural South, and thus were probably small. It therefore seems likely that less than 7.4 percent of the Talent schools were all black. Conversely, racially mixed schools were most likely to be found in the urban North in 1960 and were probably larger than the national average. One would therefore expect more than 3.2 percent of all Talent high schools to be between 10- and 49-percent black. Based on Talent's sampling ratios and some crude assumptions about the average size of schools of varying racial composition, the Talent schools might be distributed roughly as shown in Table 13.2.

Table 13.2

ESTIMATED DISTRIBUTION OF PROJECT TALENT SCHOOLS

Percent Black	Number of Schools
None	650
0-9	175
10-19	30
20-29	15
All	50
Total	920

Assuming the initial sample was racially representative of the ethnic mix of 14- to 18-year-olds in the United States in 1960, it should have included something like 40,000 black students. Since the all-black schools were probably smaller on the average than the racially mixed schools, the distribution of black pupils would not have been quite comparable to the distribution of schools shown above. If the all-black schools were half as large as the average American high school (if they averaged 200 ninth to twelfth graders), they would have enrolled a total of about 8000 blacks. The other 32,000 would have been in schools with some whites. However, this probably overestimates the actual degree of desegregation in 1960. Allowing for a few big all-black northern high schools, and for a few big all-black southern urban high schools, there might have been more like 10,000 to 12,000 black students in all-black high schools participating in Project Talent.

The second major problem with an analysis of Project Talent data is that the follow-up surveys which contain racial information have very low response rates. Talent conducted follow-ups one, five, and eleven years after students would normally have been expected to finish high schools. The 11-year follow-up of eleventh and twelfth graders in 1960 is thus complete, that of tenth graders is virtually complete, and that of ninth graders will begin in September 1974.

The initial five-year mail follow-up of twelfth graders in 1965 had a 39-percent response rate. Three years later, the initial mailing to ninth graders elicited a 29-percent response rate. Four years after that, in 1972, the initial 11-year mail follow-up of eleventh graders got a 25-percent response rate. Presumably the initial 11-year follow-up rates for tenth and ninth graders will be even lower, perhaps as low as 20 percent.

The response rate for black students is even worse. Kapel, in his analysis of the data, estimated that there were 8900 black students in the original twelfth-grade sample. Yet only 1304 returned the initial five-year mail follow-up. This implies an initial response rate of 15 percent for blacks, compared to 39 percent for the total sample. There was also a strong sex bias in black (though not white) response rates. About 20 percent of black females responded, compared to only 10 percent of black males. Assuming that subsequent reductions in the initial response rate will affect blacks and whites equally, the initial response rate in the 11-year follow-up is likely to average about 6 percent for black males and 12 percent for black females. This would yield 1200 black males and 2400 black females. It might be difficult to convince a skeptic that these 3600 blacks were representative of their 36,400 missing black classmates.

Fortunately, Talent has conducted an intensive follow-up of a subsample of those who failed to return the initial mail questionnaire. This special follow-up covers 2500 students a year, which makes it a 3- or 4-percent subsample of the initial nonrespondents. Initially, Talent only succeeded in getting data from about two-thirds of these "specials," but increasingly vigorous follow-ups have raised the response rate to 84 percent for the 11-year follow-up of eleventh graders. This rate is again lower for black respondents. In the five-year follow-up of twelfth graders, for example, the response rate for black specials was only 50 percent, compared to about 65 percent for whites. If this pattern still holds, the 84-percent overall response rate for specials in the 11-year follow-up of eleventh graders implies a response rate of about 67 percent for black specials. For the eleventh grade 11-year follow-up, then, the distributions might look roughly like those shown in Table 13.3. The absolute numbers of cases should be somewhat larger for the tenth and ninth grades, but the response rates are not likely to be much different.

By subsampling the initial respondents of the five-year and 11-year follow-up at the rate selected for the special subsample and pooling these responses with those

Table 13.3

ESTIMATED DISTRIBUTIONS OF ELEVENTH-GRADE ELEVEN-YEAR FOLLOW-UP OF BLACKS

Item	Total	Estimated Blacks
Initial sample	92,000	9,200
Responded to initial mail follow-up	23,000	828
Did not respond to initial follow-up	69,000	8,372
Selected for special follow-up (p = 0.0361)	2,491	302
Located in special follow-up	2,326	250
Cooperated in special follow-up	2,104	200

cooperating with the special follow-up of initial nonrespondents, one could obtain a new sample which is likely to be fairly representative of the original population. For example, a 3.61 percent subsample of the 828 blacks who responded initially would consist of a new sample of 30 responses. These could be pooled with the 200 respondents who cooperated with the special follow-up to produce a new sample of 230 respondents.[1] Construction of a comparable sample for whites could be achieved by following a similar procedure which oversampled those whites in desegregated schools. The major problem with this approach is that the absolute numbers it yields for black respondents are very small.

A modification of this approach would increase the absolute numbers by using *all* the initial respondents and weighting the special respondents up to approximate their true proportion. This approach does not solve the problem of small absolute numbers, however, because the standard errors for this sample will still have to be computed on the original 200 specials. What the approach does, however, is provide a mechanism for including all of the variation present in the initial responses.

Another approach would be to pool the responses from all four grades, sampling 3.61 percent of these and adding them to the special respondents from all four grades. With this approach, the numbers would be greater but the progressive effects of improved education and income for black respondents would be hidden by lumping them during a crucial period (1965-70 and 1971-75). Because of the rapid civil rights-related policy changes during the past 12 years, it might be interesting to keep the grade levels separate in the hope of identifying policy impacts for particular age groups.

Construction of a comparable sample for white students in order to assess the effect of racial composition is not fraught with the same difficulties because the

[1] The differential response rates for black males and females will present a problem if the analysis focuses on males, reasoning that male income and education are more important determinants of family SES than female. We regard this argument as specious, partly because the SES of black families is determined by the income and education of workers regardless of sex, and also because the tendencies of this post-industrial society (e.g., more leisure time, smaller families, more labor-saving devices, and increasing pressures on industry for Affirmative Action in hiring and promotion) are rapidly making the male-dominant family an anachronism. It is to be hoped that this study will not serve to perpetuate the sex bias found in much policy-oriented research. We therefore recommend that the data be analyzed with this in mind.

absolute numbers are much larger. Analysis of all the 360,000 cases would be expensive and unnecessary for our purposes. Since we are especially interested in the effects of desegregation, we should construct a subsample which overrepresents racially mixed schools and underrepresents segregated schools.

Based on the estimates of Project Talent's size and racial composition, the sample of white students might look like that shown in Table 13.4. If money were tight, the samples for the all-white and 99- to 90-percent white schools could be cut in half.

Table 13.4

SUGGESTED SAMPLING OF PROJECT TALENT WHITES

	Percent White				
	All	99.9-90.9	90-81	80-1	Total
	Number of Schools				
	650	175	30	15	870
Estimated white students	220,000	120,000	15,000	5,000	360,000
Estimated whites returning mail questionnaire	55,000	30,000	3,750	1,250	90,000
Estimated whites returning data in special follow-up	5,500	3,000	375	125	9,000
Suggested sample of whites with mail questionnaires	1,017[a]	1,110[a]	1,250	1,250	---
Suggested sample of whites with special follow-up data	2,250	3,000	375	125	---
Total suggested sample	3,267	4,110	1,625	1,375	10,377

[a] 3.6 percent of line 3. Since the special follow-up is presumed to include 3.6 percent of the initial nonrespondents, a 3.6-percent sample of the initial respondents yields an overall sample that is both self-weighting and reasonably representative.

YOUTH IN TRANSITION

The Youth in Transition study began in the fall of 1966 with a national probability sample of 2213 tenth-grade boys in 87 schools (Bachman et al. 1967, 1971). This panel was followed each year until the summer of 1970, one year after most of them finished high school, with complete data being obtained for 73 percent of the original panel of 1620 boys. Also, educational status was obtained for 376 boys who did not participate in the 1970 wave, bringing the response rate to 90 percent for that variable.

There were 256 black students who began the study in 1966. About one-half were in southern segregated schools, one-quarter were in northern segregated schools, and one-quarter in northern desegregated schools (defined as 40 percent or more white students). The response rates were somewhat lower for black students, particularly for northern desegregated students; complete data on educational status was obtained for about 80 percent of this group compared with 90 percent for white students. This means that there is complete follow-up data on educational status for about 50 black students who attended northern desegregated high schools.

While these numbers are quite small compared to Project Talent, the data is richer in post-high school outcome variables. The study not only included the usual questions about career patterns—such as college attendance—but a number of racial attitude measures as well. These attitudes were assessed in both the 1969 and the 1970 waves, so that there is a base-line measurement in the senior year of high school and another measurement one year after high school. Attitudes measured included endorsement of strong government action to insure racial equality, perceived discrimination of blacks, and a standard social distance (or contact) scale.

It appears, therefore, that the Youth in Transition data are yet another resource for examining differences between black and white students, both segregated and desegregated, on a number of outcomes, including college attendance and racial attitudes one year out of high school. Moreover, it would probably be feasible to reinterview this panel again in a new study focused on desegregation effects; at this time it would be possible to include measures of those variables shown to be important by the main desegregation study.

NATIONAL LONGITUDINAL STUDY

The National Longitudinal Study, sponsored by USOE, began in 1972 with a national probability sample of about 20,000 students in approximately 1200 schools (Berry 1973).[2] As we pointed out in Section 3, this study is just under way and there is only preliminary information on results of the first follow-up year (Chandler 1974). The study design calls for the use of the U.S. census interviewing staff to obtain a 90-percent response rate; about 60 percent are expected to respond to an initial mail questionnaire.

Given these numbers we would expect about 2000 black students in the original sample and smaller numbers of other minority students. The school questionnaire included racial composition questions pertaining to both students and staff, so it should be possible to group these students into segregated and desegregated schools. At the present time, however, it is not known what the numbers will be.

The follow-up questionnaire focused primarily on career pattern questions, including college attendance, type of job, reasons for choosing college or the job, and so forth. Thus the data should be valuable for comparing segregated and desegregated minority and white students with regard to career patterns, including college attendance.

Since USOE plans to continue reassessments for up to six or eight years, it may be possible to include race-relations questions in some future follow-up wave. Promising variables could be proposed on the basis of the first-year results of the main desegregation study. If so, the National Longitudinal Study would become an extremely valuable (and inexpensive) resource for studying differences between segregated and desegregated schools with respect to a number of post-high school outcomes.

These three studies provide the opportunity for a cost-effective study of the postschool effects of desegregation. Such studies are valuable. Should they fail to provide clear answers to the research questions, or, as is more likely, should they advance our knowledge to the point where we can ask more complex questions about the impact of public school race relations on postschool behavior, then a new study can be undertaken, using the twelfth-grade students of the original large-scale longitudinal survey as the beginning of a new panel.

[2] Some of these students did not participate in the first wave but students from them were included in follow-up waves.

14. SOME SUGGESTIONS FOR EXPERIMENTAL RESEARCH ON FACTORS INFLUENCING ACADEMIC ACHIEVEMENT IN DESEGREGATED SCHOOLS

PREFATORY COMMENT

Most research to date on school desegregation has used surveys and questionnaires almost exclusively. In part because of the limitation of this method, results have not been as clear as we had hoped. We have proposed in this report that quasi-experimental and experimental design, longitudinal surveys using better instruments, and observational case studies, be combined to provide better data on the subject. But even this wide range of techniques is inadequate, because it will not provide an opportunity to develop theories of the social-psychological dynamics of the learning experience in a desegregated school. We have asked Irwin Katz to draw upon his own experiments in social psychological research in race relations and prepare recommendations for further laboratory work.

He suggests several research projects; it is obvious that there are many more. The amount of laboratory work that can and should be done is quite large. For example, one might develop a large number of studies in which the subjects are public school teachers and administrators, manipulating factors which we hypothesize would affect their behavior toward minority students. If such studies help us develop a theory of the teacher's reaction to desegregation, our chances of developing effective desegregation plans are greatly increased. Our conclusion is then to recognize that the laboratory experiment is our most powerful tool for testing social psychology theory, and that a number of experiments, including those decribed by Katz, are needed to supplement the surveys, social experiments, and case studies described in this report.

INTRODUCTION

Perhaps the more important reason why previous research has not been able to unequivocally determine the effects of school racial composition on minority performance is that very few studies have tried to identify the specific pupil characteristics and in-school factors that determine whether a given desegregated situation will have a favorable, unfavorable, or neutral impact on children's intellectual growth. Yet this appears to be the research focus that is most likely to yield useful knowledge for improving the educational effectiveness of racially integrated schools. Both practical educational experience and psychological research have generated a number of interesting questions and hypotheses about the motivational determinants of minority children's performance and learning in racially mixed settings. These have to do, for example, with the influence of (a) own-race and other-race age peers as behavioral models, setters of achievement standards, and dispensers of positive and negative reinforcements (approval and disapproval); (b) the racial identity, attitudes, expectancies, and behavior of teachers; and (c) pupils' expectancies (subjective probabilities) of success and failure, and self-evaluations of performance (including causal attributions of success and failure).

A few attempts have been made in large-scale desegregation studies to evaluate the role of attitudinal, motivational, and social-interactional factors in minority-

group children's academic achievement (Coleman et al. 1966; U.S. Commission on Civil Rights 1967). However, such surveys must rely almost exclusively on correlational analyses of questionnaire and standard achievement test data, and hence they are not well suited for systematic investigation of causal relationships. To establish causal relationships it is highly desirable to augment the survey findings with data from well-controlled, laboratory-type experiments. This section will describe some experimental paradigms that give promise of helping to unravel the nonintellectual determinants of minority pupils' scholastic performance in desegregated classrooms.

FACTORS MODIFYING THE IMPACT OF BIRACIAL SETTINGS ON MINORITY-STUDENT PERFORMANCE OF MENTAL TASKS

1. Influence of Teacher-Examiner's Race and Evaluative Significance of the Task. Some years ago, Katz, Roberts, and Robinson (1965) did an experiment in which the subjects were students at a predominantly black college in the South. Subjects were individually administered a simple mental task (digit-symbol substitution) by either a black or white adult, with instructions which defined the task either as a test of intelligence (IQ) or as a test of eye-hand coordination. It was found that when the task was called an IQ test, performance was slightly better with a black examiner than with a white examiner, but when the identical task was presented as a nonintellectual test, a sharp reversal occurred; performance was substantially better with the white tester. The white-examiner/motor test condition occasioned better performance than any of the three other conditions. The results were interpreted in terms of an incentive and probability model. This entailed the following assumptions: (a) that the incentive value of success was greater with the white than with the black examiner, and (b) that the subjective probability of success was lower with the white than with the black examiner under IQ test instructions (attribution of higher intellectual achievement standards to the white than to the black examiner), but about the same with both examiners under motor test instructions. Subsequent research has tended to support these assumptions when applied to black college students.

This experiment has not been replicated with black subjects of elementary and high school age, yet the paradigm seems well suited for exploring the effect of evaluative *versus* nonevaluative teacher orientations on the responsiveness of black pupils to teachers of varying racial identity in academic achievement situations. Of particular interest is the implication that under certain conditions merely placing minority children with teachers of the dominant race (or raising the status and prestige of own-race teachers) may have a beneficial effect on classroom performance. Atkinson's (1964) theory of achievement motivation suggests that the favorable conditions for cross-racial matching of minority pupils and teachers are those in which success is seen by the child as a more likely outcome than failure, and/or the evaluative significance of failure is minimal.

Other independent variables that could be built into replications of this paradigm include the ability level of pupils, sex of pupils and examiners, prior experience of pupils with white teachers, difficulty and qualitative features of the learning task.

2. Influence of Race and Peer Comparison Group and Race of Examiner. In another experiment (Katz et al. 1972) mental tasks which were described as measures of academic ability were given to freshmen at a predominantly black college in the South. The tasks were administered by either a black or white adult

experimenter who announced that scores would be evaluated by comparing them with the average scores of either white age peers or black age peers. It was assumed that the black subjects would tend to perceive the white testers as more authoritative (i.e., more powerful and prestigious) evaluators than the black testers, and that the comparison with white norms would be seen as a more difficult achievement situation than the comparison with black norms. Hence it was predicted that in the white norm (i.e., relatively high risk of failure) condition performance would be better with a black (i.e., low status) examiner, whereas in the black norm (i.e., relatively low risk of failure) condition performance would be better with a white (i.e., high status) examiner. This prediction was confirmed by the results. Moreover, subjects' self reports of their perceptions on a post-experimental questionnaire tended to support the two assumptions underlying the prediction.

Another independent variable in this study was the probability of success. As part of their task instructions, subjects were told that the scores they had attained at a prior practice session indicated how likely they were to at least equal the norms with which they were to be compared. They were randomly assigned to three probability-of-success conditions: very low probability (10-percent chance of success), moderate probability (60 percent), and very high probability (90 percent). It was predicted that as probability of success increased, overall performance with the white examiners would improve relative to overall performance with the black examiners, so that in the 90-percent condition the white examiner group would be clearly superior. The results showed only a weak tendency in this direction. However, a replication of this experiment which employed a more credible manipulation of probability of success confirmed the prediction.

The foregoing research indicates that with black subjects of college age, at least in predominantly black colleges in the South a few years ago, the effects of cross-racial competition, cross-racial evaluation, and objective probability of success on achievement motivation were complex and interdependent. Moreover, when success seemed attainable, biracial achievement situations were more intensely motivating than all-black situations. It remains to be ascertained whether these conclusions are applicable to black pupils of elementary and high school age in various regions of the country today. The experimental paradigm is a relatively simple and economical one to execute, and gives promise of providing basic insights into social comparison processes and reactions to adult authority figures in racially mixed achievement settings. As in the case of the first paradigm discussed in this section, a whole array of variables can be plugged into it, including the type of tasks employed, the level of academic ability of minority group subjects, age and sex of subjects and peer comparison groups, amount and quality of prior academic competition and social interaction with white peers, prior experience with white teachers, and so on. Such experiments would provide a means of exploring the suggestions of Pettigrew (1969) and Gerard and Miller (1971) regarding the motivational importance of cross-racial peer comparisons in the desegregated classroom.

SOME DETERMINANTS AND BEHAVIORAL EFFECTS OF MINORITY-PUPIL SELF-EVALUATIONS OF PERFORMANCE

Previous studies (Coleman et al. 1966) clearly document the academic difficulties of minority-group children. Yet little is known about the motivational aspects of this problem. Elsewhere the theory has been proposed that the persistence required for academic success depends heavily upon an internalized mechanism consisting of

affect-mediating self-evaluations (Katz 1967). Academically successful individuals are presumed to sustain their efforts toward academic goals by administering covert self-evaluations of their own strivings. Favorable self-evaluations are experienced as rewarding (have positive reinforcement value), whereas unfavorable self-evaluations are experienced as punitive (have negative reinforcement value). In these terms, the difficulties of academically unsuccessful black pupils may stem from dysfunctional self-evaluative processes elicited in achievement-related situations.

In research conducted to explore this perspective (Katz 1967), fourth- through sixth-grade black children were taken individually from their classrooms for testing. During a self-evaluation phase, a series of simple tasks (picture assembly or construction of four-letter words) was presented to each child, who was seated alone at a table and surrounded by partitions. Near the child on the table was a metal box with three buttons which activated small light bulbs of different colors labeled "good," "poor," and "don't know."

The instructions were in part as follows: "We think you will enjoy doing these things more if you can tell yourself how nice a job you think you did. So after you finish each one you can press the button which shows how you feel about the kind of job you did.... No one will know which button you pressed." The experimenter left the room after explaining the procedure. Hence the self-evaluations were ostensibly private, unobserved, and for the child's own amusement. But the setup was deceptive: the button pressings were mechanically recorded by counters concealed in the box.

In another phase of the testing, the extent to which the child's self-evaluations had reinforcement properties was investigated by ascertaining whether the colored lights used in the self-evaluation box had acquired positive or negative incentive value by virtue of being associated with self-criticism or self-approval. This was done by giving the child a coloring task with crayon before and after the self-evaluation phase. Changes in amount of use of the colors that were associated with the "good" and "poor" buttons were supposed to indicate changes in their attractive or aversive properties.

It was found that academically unsuccessful boys were more self-critical and less positive in overall self-evaluations of their performance on these simple tasks than were academically successful boys—despite no objectively apparent differences in the judged quality of their products. Moreover, after the self-evaluation phase the former group, but not the latter group, tended to avoid the crayon color that had been associated with the "poor" button.

In addition, questionnaire measures of perceived past social reinforcement from parents were systematically associated with subjects' task-related self-evaluations: the greater the extent of perceived disapproval and the less the perceived approval from parents, the less favorable the boys' self-evaluations of their products. Further, academically unsuccessful boys recalled less approval and more disapproval from their parents than did high-achieving, academically successful boys.

Extrapolating from these findings, one may suggest that a history of predominantly negative social reinforcement from parents and/or teachers may underlie the low-achieving black youth's penchant for overly harsh and self-defeating criticisms of his own accomplishments—a potential key factor behind his academic failure.

Recently, Dion and Miller (1973) explored this theory further. Black children's self-evaluations of their performance on an ambiguous task were assessed as a function of experimental variations in prior social reinforcement and subjective privacy of self-evaluation. In the private self-evaluation condition, subjects given disapproval-oriented reinforcement were less self-approving and less positive over-

all in evaluating their own performance than those receiving approval-oriented reinforcement. Moreover, they exhibited lower levels of aspiration on a subsequent ring-toss task. In contrast, subjects whose self-evaluations were public rather than private failed to exhibit these effects.

Proposed Experiments. The research just described demonstrates the feasibility and potential value of studying minority children's covert self-evaluations of their performance in academic situations. If indeed dysfunctional habits of self-evaluation are impairing the low-achieving black child's will to learn, it would be highly desirable to investigate further how these habits are acquired and how they can be modified. A number of possible lines of inquiry suggest themselves, using the experimental techniques described above.

First, both Katz (1967) and Dion and Miller (1973) used only white male experimenters and black male subjects. Further research should vary the race, sex, and social class characteristics of both subjects and experimenters. Dion and Miller's (1973) study dealing with the effect of adult approval or disapproval on children's self-evaluations should be repeated using these variations in subject and adult characteristics.

Second, various experimental paradigms of social learning that have recently been developed could readily be adapted for the study of social influences on children's self-evaluations. Bandura and Kupers (1964), Bandura and Whalen (1966), Mischel and Liebert (1966), and others have studied children's imitation of peer and adult models' self-rewarding behavior in game-like situations. These investigators used material self-reward (e.g., the taking of candy from a mechanical dispenser), but private self-evaluation could readily be substituted as the dependent variable to ascertain social influence effects of models with various characteristics.

Such experiments would establish whether exposure to the self-evaluative behavior of peer and adult models can influence (a) the favorableness of the subject's self-evaluations, and (b) the degree to which self-evaluations mediate positive and negative affect. Once the imitation effects had been demonstrated, one could investigate experimentally the modifying influence of certain variables; e.g., the degree of discrepancy between the models' competence and the subjects' competence on the task being performed, the degree of discrepancy between models' and subjects' evaluative standards, the race of models, the reinforcement history of subjects in achievement situations involving similar tasks, and the general level of the subjects' self-evaluative standards on similar tasks. Several questions of theoretical interest are relevant to the proposed experiments. For example, there is the issue of whether the low academic achievers, who were more self-critical in this writer's pilot study than high achievers, have stronger or weaker internal standards than the latter. Since the more strongly internalized standards should be less susceptible to social influence, the extent to which high and low academic achievers shifted their initial standards on a task in the direction of the standards of models would bear directly on the issue.

Third, the motivational consequences of public and private self-evaluations should be explored. Dion and Miller found that only private self-evaluations influenced motivation (level of aspiration) on a second task. Further research should employ this measure as well as additional measures of motivation such as intensity of performance and persistence of striving.

The foregoing discussion of directions for research on self-evaluation is not intended to be exhaustive, but merely to indicate the range of possible issues that can be investigated, with respect to the etiology, modification, and emotional effects of children's self-evaluations of their performance on various school-related tasks.

MINORITY CHILDREN'S PERCEPTIONS OF THE CAUSES OF THEIR SUCCESSES AND FAILURES

Recently, Weiner et al. (1971) have reported on a program of research into the causes that children and teachers ascribe to their own and others' successes and failures in achievement situations. They postulate that individuals utilize four elements of ascription both to postdict (interpret) and to predict the outcome of an achievement-related event. The four causal elements are ability, effort, task difficulty, and luck. That is, in attempting to explain the prior outcome (success or failure) of an achievement-related event, the individual assesses his own or another performer's ability level, the amount of effort expended, the difficulty of the task, and the magnitude and direction of experienced luck. Similarly, future expectations of success and failure are based upon the assumed level of ability in relation to perceived task difficulty, as well as an estimation of intended effort and anticipated luck.

It can be seen that two of the four components in the model (ability and effort) describe qualities of the person undertaking the activity, while the two remaining components (task difficulty and luck) are regarded as properties external to the person, or environmental factors. Further, two of the elements (ability and task difficulty) have somewhat enduring characteristics, whereas the magnitude of the two other components (effort and luck) are relatively variable. Thus, Weiner and his associates refer to two basic dimensions: locus of control (internal versus external) and degree of stability (fixed versus variable). In their research they have explored the antecedent conditions affecting causal attributions, and the effects of various types of causal attribution on subsequent behavior, such as self-reward and intensity of achievement striving.

Using measures of individual differences in achievement orientation similar to those used by Atkinson and others (1964), Weiner's group have studied white school children and college students. Following are some of their relevant findings.

Given *success* at a task, subjects high in achievement orientation are more likely to attribute the outcome to internal factors (ability and effort) than are subjects in the low achievement-motive grouping, while subjects in the low motive group are more prone to ascribe success to the external factor of task ease. Further, individuals high in achievement orientation reward themselves more (e.g., take more poker chips from a bowl) for success than do individuals who are low in achievement motivation, and the greater the tendency to believe that success, but not failure, was caused by oneself, the greater the resultant self-rewarding behavior relative to self-punishing behavior (returning poker chips to the bowl).

The data also indicate that in situations of *failure,* individuals high in achievement motivation attribute their poor performance to lack of effort, while individuals low in achievement motivation perceive failure as being due to insufficient ability. Finally, attributions of failure to bad luck and lack of effort result in better subsequent performance than attributions to lack of ability or task difficulty.

Proposed Research. The work of Weiner et al. (1971) on children's causal attributions seems to demonstrate that (a) these attributions have important motivational properties, and (b) children classified as high or low on achievement orientation, as defined by Atkinson and others, display strikingly different attributional tendencies. The high achievement-motive children appear to have acquired attributional habits that *enhance* their capacities for classroom learning and performance, whereas low achievement-motive children seem to have developed attributional habits that *impair* these capacities.

Up till now, virtually all of the attributional research has been done with white male subjects. It would be desirable to ascertain whether underachieving black

children display the same self-defeating attributional tendencies as do the Weiner group's low achievement-oriented white male children. Katz's (1967) research on black pupils' self-evaluations would suggest this hypothesis. A first line of inquiry might be to replicate the Weiner group's studies using minority-group children from various social class backgrounds. If it is found that underachieving black pupils are indeed prone to make self-defeating *attributional errors* (i.e., to ascribe their successes to task ease and luck, and their failures to lack of ability), then further research should focus on the causes of these false and dysfunctional attributions, and on techniques for corrective attributional training. Of interest would be the role of teacher attributions regarding the causes of pupil underachievement. For example, the Weiner group have suggested that teacher attitudes toward children with learning difficulties may be a key factor in the development of self-defeating or self-enhancing attributional tendencies. Teachers who have negatively biased expectations about the achievement of black pupils may give off interpretive cues that form the basis of the child's erroneous attributions.

One could investigate the effectiveness of sensitizing children to relevant internal processes so that, for example, they could recognize when they are trying to do well on a task and when they are being distracted by fear of failing. An adult model could provide *verbal labels* for these and other internal states that influence his performance. The training should be concerned as well with improving the child's ability to use information from the environment, such as the knowledge of the relative difficulty of tasks that can be gleaned from social comparisons with age peers. Enlarging the child's perceptual field for attributions will require in some instances reducing the affective consequences of success and failure, so that he or she will not be overwhelmed by the salience of his or her own behavior.

These are just a few suggestions for research on minority pupils' causal attributions of their successes and failures. Once it has been demonstrated that *attributional errors* are a key factor in their achievement difficulties, a variety of useful research directions will become apparent.

SOME CONCLUDING REMARKS

The foregoing discussion of possibly useful and theoretically interesting types of research on minority-student motivation for academic achievement is not in any sense intended to be exhaustive. There are numerous other potentially valuable types of inquiry that have not been mentioned; for example, research on heterogeneous versus homogeneous ability grouping in classrooms, variations in the racial-composition ratios of pupils, manipulation of evaluative versus nonevaluative teaching styles, direct observations of the behavior of teachers whose expectation about pupil achievement have been systematically manipulated (i.e., follow-ups on Rosenthal and Jacobson's "Pygmalion in the Classroom" effect). In addition, controlled experiments might profitably include a range of personality variables, such as school anxiety, internal-external fate control, independence versus conformity tendencies, and self-esteem.

The feasibility of this research is reflected in the large number of able experimentalists whose interests and qualifications would make them appropriate for this sort of work. The names listed in the bibliography are only a small sampling of such people.

15. DESIGN ALTERNATIVES AND COSTS

In this section, we consider what opportunities exist to modify the study and to reduce its costs, either altogether or in the initial year. We list the separate tasks of the study, attach approximate cost estimates, and then consider alternatives in either the scheduling of the tasks or the alteration of the study to eliminate certain tasks.

The total study, as described in this report, requires six years and $9 million to $15 million. Is such a large study justified? We think so. If 30 years of small and uncoordinated studies have taught us little about school desegregation, their proposing another small study or a few uncoordinated studies is not helpful. A large, tightly coordinated program is needed because policy-makers have a very wide range of interests related to school desegregation and because science advances through coooperation and competition among scientists with differing hypotheses and differing methodologies.

Nevertheless, we recognize that budgets are finite, and that other policy issues compete with desegregation for research funds.

To assist the reader in making decisions about funding, the following pages review the study tasks and attach cost figures to each.

OUTLINE OF TASKS INVOLVED IN THE RESEARCH PROGRAM

Table 15.1 lists eighteen separate research tasks that make up the research program and a very approximate cost estimate for each. Table 15.2 indicates a time schedule for these tasks. In this section we briefly describe each task, highlighting the research capabilities that are most important. In general, we recommend that decisions regarding contract awards for these tasks be based heavily on corporate and research capability, as long as such an emphasis does not prejudice the funding process against young researchers and researchers without previous experience in desegregation. Too few researchers have experience in desegregation; one of the opportunities this program presents is to increase the number of people involved in the field.

The *management* task involves assisting the contract monitor in maintaining the study schedule. The responsible researcher should be one with organizational skills, an ability to establish good working relationships with clients and contractors, considerable research experience, and a willingness to remain involved in the project for a long period of time.

We recommend that all *data collection* tasks and preparation of data tapes be done by a single contractor. Economies of scale here are quite important. The contractor should have (1) experience in surveys of the national population, surveys of community leaders, and data collection from schools; (2) capability for optical scanning of questionnaires and experience preparing data tapes aggregated to various levels of analysis; (3) a good record of delivery of client-usable data tapes on schedule.

The data collection task is large but not unduly so; for example, the total amount of data collected is less than that collected in *Equality of Educational Opportunity.* Data to be collected are summarized in Table 15.3 which shows the number of

Table 15.1

LISTING OF MAJOR TASKS IN PROGRAM

	Task	Costs ($ million)
1	Management	0.2-0.4
2	Data collection, preparing data tapes	5.8-8.1
3	The effects of district desegregation on parental and community reaction	0.2-0.4
4	Describing what parents want from schools, their feelings about desegregation	0.05-0.1
5	The effects of the school: a cross-sectional analysis of the first-year data	0.15-0.25
6	Selecting samples for substudies	0.05
7	The effects of the schools on students and their parents: analysis of large-scale longitudinal study	0.25-0.35
8	The effects of the school: longer-term effects on students from panel survey	0.2-0.3
9	The effects of the school on Spanish-language students and their parents (from large-scale and panel surveys)	0.3-0.4
10	Describing the operation of newly desegregated schools (including case studies)	0.6-1.0
11	Evaluating elementary school innovations from process substudy	0.15-0.3
12	Evaluating high school innovations from process substudy	0.15-0.3
13	The effects of desegregation on suburban migration	0.05-1.0
14	Secondary analysis of Project Talent, national longitudinal study, youth in transition	0.1-0.2
15	Experimental social-psychological research	0.3-0.7
16	The effects of racial experiences (including school desegregation) in childhood on the present generation of adults (from parent survey)	0.05-0.1
17	Unsolicited additional analyses, case studies, etc.	0.25-1.2
18	Synthesis of results for policy process	0.15-0.3
		9.0-15.45

Table 15.2

SCHEDULE OF MAJOR TASKS

Fiscal Year, Month	District Studies	Large-Scale Longitudinal Study	Student Panel Survey	Study of Newly Desegregated Schools	Elementary and Secondary Innovations	Other Studies
1 July August September October November December January February March April May June	Contract awarded Screener sample District sampling District leader survey[a] Parent group survey[a]	Contracts awarded for analysis, data collection School sampling OMB (a)		Contract awarded		1 Award contract for secondary analysis of existing data on post-school behavior
2 July August September October November December January February March April May June	Parent survey[a] Second screener survey[a] Electron results[a] Report, parent values about education	School data tape complete School parent 1 year tape (a)	Contract awarded Sampling Modify quex?	Sampling Contract awarded Mobility quex? Begin case study		Award contract for *Resegregation Study* Award contract for laboratory experiments Reports at intervals for rest of project
3 July August September October November December January February March April May June	Final report district desegregation study Rpt. effects of exper. in deseg. schs. on parents Parents in desegregating districts[a]	Data tape completed Final report, first year analysis	(a)	(a)	Decide on E and S innovations Sampling Evaluation analysis Contract awarded (a)	From post-school secondary analyses recommendation on need for a 12th to post-high school survey
4 July August September October November December January February March April May June	Interim report Parent reaction	Final report, longitudinal analysis	Interim report, Panel survey (a)	(a)	Third year tape Revise quex (a)	Resegregation study final report Final report, post-school secondary analyses
5 July August September October November December January February March April May June	Parents in desegregating districts Parents in desegregating districts[a]		Final Report, Panel survey	Interim report End case study fieldwork Parents resurvey (a)	Fourth year tape Revise quex (a) Interim report	Synthesis of laboratory studies
6 July August September October November December January February March April May June	Final report, parent reaction			Final tape Final report	Final tape Final report	
	---------- POLICY RECOMMENDATIONS ----------					

[a]Data collection point (for school survey unless otherwise indicated).

Table 15.3

APPROXIMATE NUMBER OF QUESTIONNAIRES ADMINISTERED[a] IN PROGRAM BY RESPONDENT AND YEAR

Year	Studies	Number of Questionnaires					
		Community Leaders	Parents	Schools[b]	Students	Teacher	Principals
1	Screener, community reaction, large-scale school study	2,250 2,960	1,581	1,623	165,000	23,000	1,623
2	Community reaction, large-scale school survey, student panel parents, first grade student panel		15,000	1,623	170,000	23,000	1,623
3	Screener, newly desegregated schools, student panel, innovative elementary and secondary schools, follow-up of parents in desegregating districts	2,250	1,000	538[c]	75,000[c]	7,250[c]	502[c]
4	Newly desegregated schools, student panel, innovative schools		5,000	616[c]	95,000[c]	7,000[c]	580
5	Screener, newly desegregated schools, innovative schools, resurvey district leaders and parents in desegregating districts	2,250 850	2,000	385[c]	35,000[c]	6,000[c]	385
	Total (approx.)	10,500	25,000	1,623	540,000	64,000	4,500

[a]Telephone, mail, personal interviews.

[b]Assuming a 6-3-3 grade structure.

[c]Assuming overlap in substudies at a random rate.

questionnaires administered per year. (A more detailed summary is given in Appendix A.)

The effects of district desegregation on parental and community reaction is a problem in empirical political science or political sociology. The contractor should be able to use effectively the literature on community decisionmaking and the relationships between voters and political leaders. The most difficult part of this project is the location of historical events in time from the survey so as to permit cause-and-effect arguments.

Describing parental values and attitudes toward school is a task in traditional sociological research, but is an extraordinarily difficult one. The survey instrument for this part of the study was the most difficult to write, and the contractor should have the capability for making improvements. Analysis involves careful attention to response bias and the ability to interpret the meanings of survey responses with considerable wisdom.

School effects from the first year data. This task is likely to be done poorly simply because the opportunities to do longitudinal research will cause us to underestimate

what can be learned from the first wave of data. Analysis of these data is important since they will be the first school-related results in the study and can heavily influence revision of questionnaires for later administrations. The contractor should have the ability to work with regression techniques, and be aware of problems in reliability and sources of error in multiple regression. The contractor should be able to deliver results promptly.

Selection of samples for substudies involves the relatively simple task of verifying that the matched pairs are indeed satisfactory matches. The contractor should be able to use the literature on bias in matched samples.

Analysis of the large-scale longitudinal study is one of the most important tasks in the project. This study has the capability of providing tests of an extremely wide range of hypotheses and should result in a report which in scope and depth far exceeds that of *Equality of Educational Opportunity.* The contractor should be able to handle a very wide range of hypotheses, and be able to use the very extensive literature on school and teacher effects, race relations, and the effects of desegregation. The contractor should be able to handle noncognitive outcomes well. This task has the largest budget of any of the analysis tasks in the program.

Analysis of the student panel data requires a high level of skill in handling the methodological problems of longitudinal data. The contractor should be able to analyze noncognitive as well as cognitive data in a longitudinal design, with particular attention paid to development of growth curves, measurement of retention characteristics through childhood and adolescence, and establishing causal relations among student variables through techniques such as cross-lagged panel designs.

The analysis of Spanish-language students and their parents may be understudied because most of the potential contractors will be more interested in the analysis of black-white relationships. The contractor should have the ability to analyze the school experience of Spanish-language students in the light of linguistic and cultural factors. At the same time, the contractor needs reasonably strong methodological skills, will need to handle both the large scale and student panel surveys, and will need to be able to analyze a wide-ranging set of hypotheses. If no single contractor has the necessary methodological experience, it may be necessary to develop a cooperative relationship between this contractor and the contractors analyzing the large-scale longitudinal study and the student panel.

Describing the operation of *newly desegregated schools* requires a contractor who is capable of making a smooth connection between the collection and analysis of case study materials and the use of survey data. We caution the reader that this combination of skills is extremely rare. The contractor should be able to use the literature on race relations in schools and have a good awareness of school operations.

Evaluating elementary school innovations requires a contractor with unusually good methodological skills combined with a strong ability to understand the nuances of educational innovations.

The evaluation of high school innovations is similar to the elementary school project, but requires more attention to problems of adolescents and problems of peer group relations.

The effects of desegregation on suburban migration is a methodologically difficult project which requires good use of census data, a good understanding of problems of multicollinearity in regression, and a high level of objectivity. The contractor should be able to use theories of housing movement.

Secondary analysis of existing data on post-high school behavior requires reasonably good methodological skills, plus the ability to use data on occupational behavior.

Experimental social psychological research requires research teams well trained in traditional laboratory techniques, but able to use hypotheses from personality theory and education as well as social psychology.

The analysis of parent data to assess the impact of school racial experiences on adult behavior is an interesting example of the economies of scale in the research program; the same research project, funded separately, would cost perhaps five times as much as it does as a by-product of the parent interviews. The research project uses conventional survey research analysis skills coupled with a strong ability to analyze sources of error in retrospective cross-sectional data.

Unsolicited proposals can be used to supplement the overall research project in a variety of ways. In general these should be for either secondary analysis from the data bank or proposals to add small amounts of case study data and other data to the data bank analysis. We anticipate that these projects will be very cost-effective because of the use of the data bank. Unsolicited proposals are an excellent opportunity to fund the testing of interesting new hypotheses and to involve a researcher who brings a particularly valuable skill to the project. However, researchers with particular strengths may lack the methodological capability to execute the work; in those cases, partnerships between them and some of the other research teams may be extremely valuable.

Synthesizing the results for the policy process is the single most important step in the program. The contractor here should have a rare combination of objectivity and courage. Too often objectivity is used as an excuse for the unwillingness to admit to any positive findings from research. But the refusal to accept hypotheses as being supported by the data, and the refusal to draw policy conclusions from data are biases as serious as simple anti-busing or pro-busing prejudice. The contractor should have a good knowledge of the policy process and a strong interest in influencing policy, coupled with good critical and analytical skills.

COSTS INVOLVED IN ALTERING THE TIMETABLE

The timetable devised for this study is the shortest one possible. Six years is a brief period in which to expect scientific research to lead to policy conclusions, and we are quite convinced that the timetable cannot be shortened. It can, however be lengthened in order to reduce the total funds needed in any one year. There are two ways this can be done. First, the community reaction survey and the large-scale longitudinal survey of schools could be begun in different years, so that no one year's budget would need to include these two large costs. Second, the study of school innovations and the study of newly desegregated schools would be delayed one or two years.

The community reaction survey is not as closely linked to the remainder of the program as other parts are. While there are good reasons for executing this study at the beginning of the program, it would certainly be possible to delay it one year. To do so does not affect total costs, but it does make the first-year budget more manageable.

It would also be possible to delay the initiation of the special studies of newly desegregated schools and innovative schools by as much as two years. This has both advantages and disadvantages. It permits more time to design these studies, provides time to carefully mine the large-scale longitudinal survey and the student panel for findings, and provides time to develop new hypotheses. The main disadvantage is that the selection of the subsamples for the studies will be damaged by the obsolescence of the large-scale survey data. This would perhaps mean that some sort

of screener resurvey of the large-scale longitudinal sample (perhaps interviews with principals) may be necessary to draw good samples. The effect of either of these two charges would be to increase total costs slightly, decrease the costs per year, improve slightly the overall study's ability to produce results, and delay obtaining policy results. The decision whether delay is wise is thus a difficult one.

OPTIONAL VARIATIONS IN THE RESEARCH PROGRAM

There remain several important decisions which funding agencies should make regarding the conduct of the research program: whether to modify the project to eliminate the study of segregated schools; whether to incorporate an experimental-design allocation of federal aid to the schools; whether to make various reductions in sampling for cost reasons; whether to add a panel of high school alumni to the sample and whether to add a more elaborate study of the adult racial behavior. Each of these alternatives is discussed below, and cost changes are shown in Table 15.4.

Eliminating Segregated Schools from the Design

At various times the staff of the Commission on Civil Rights has recommended that the comparison of segregated and desegregated schools be eliminated from the study in order to concentrate on the analysis of school process within desegregated schools. The reasons for this have been discussed in Sec. 1, Volume 1, of this report. Segregated schools are included in only two data collection efforts: the large-scale longitudinal study and the student panel survey. Eliminating segregated schools from these two studies raises no special problems and considerably simplifies the study. The large-scale longitudinal study is now built on a sample of matched quintets and quartets; dropping segregated schools would mean that it should be built on a sample of matched pairs and desegregated schools. This would reduce the number of schools from 1623 (approximately) to 780.

A similar reduction of the student panel from 80 to 40 schools in each grade level would follow. This would result in an overall saving of $1,500,000 to $1,900,000, primarily in data collection costs. We recommend that this option *not* be taken for three reasons. First, despite the fact that desegregation is legally mandated and not an issue on which policy-makers are free to make decisions based on social science data, we believe that there is a great deal of legitimate interest in the comparison of segregated to desegregated schools among policy-makers. Second, we believe that an understanding of life in desegregated schools can be achieved most easily when it is contrasted to life in segregated schools; third, we believe that segregated schools, both Anglo and minority, will remain for the foreseeable future, and that for the relatively small increase in the cost of the study, considerable information will be produced about improving the quality of education within those schools.

Modifying the Study to Incorporate Experimental Allocation of Federal Aid

We recommended in Sec. 12 above that the study be converted to an experimental test of interventions in desegregated schools, either in the process studies during the second phase of the project or in both that phase and in the earlier large-scale longitudinal study. To do so considerably improves the quality of data and results in some savings in research costs, but complicates the management of the project.

The major savings result from combination of the school process studies with the

Table 15.4

COST IMPACT OF EXPERIMENTAL MODIFICATIONS

Modifications	Cost Changes ($ million)	
	Additions	Subtractions
Process Study:		
Additions and subtractions		
1. Development of experimental interventions	0.1-0.2	
Date collection: Combine newly desegregated schools, innovations into one study, and overlap all desegregated school with panel: 125 desegregated schools + 40 segregated schools at each grade level:		-0.32-0.45
Implementation of innovations (elementary school)	3.6-6.5	
Implementation of innovations (high school)	10.7-18.7	
Evaluation of newly desegregated schools and innovations		0.5-0.9
Costs, innovations	14.3-25.2	
Net savings, research		0.7-0.12
Large-scale Study:		
Development of innovations	0.2-0.4	
Implementation of innovations in all 780 desegregated schools in large-scale study	16.4-29.2	
Evaluation of innovations: no change		
Reduction of process study by one year, implementation costs		2.5-4.6
Date collection costs		0.2-0.3
Evaluation, and other studies:		0.3-0.4
Net costs, innovations, entire program	28.2-49.8	
Net savings, research, entire program		1.0-1.7
Total project cost, process experimental version (duration 5 years)	8.3 to 14.25 + 14 to 25 in school aid	
Total project cost, full experimental version (duration 4 years)	8.0 to 13.75 + 28 to 50 in school aid	

student panel sample and the sample for the analysis of characteristics of nearly desegregated schools. Because interventions are experimentally awarded, there is no need to draw a special sample of schools that already have particular programs; thus these samples can be collapsed together. The experimental design also greatly simplifies the analysis, eliminating a large number of statistical problems. For these reasons we estimate that the use of an experimental process study would result in a saving of $500,000 to $900,000 in research costs.

Conversion of the initial large-scale longitudinal study to an experimental design permits the entire study to be shortened one year, since it eliminates the need for a one-year interlude between the large-scale study and the following process

Table 15.5

LIST OF DESIGN OPTIONS

	Costs or Savings ($ million)
Experimental process study	-(0.7-1.2)
Full experimental version	-(1.0-1.7)
Elimination of segregated schools	-(1.5-1.9)
Adding a post-high school panel	+(0.5-0.7)
Adding a study of adult racism	+(0.2-0.4)

substudies. Use of an experimental design greatly increases the efficiency of the sample, but we do not recommend that sample sizes be reduced, because cost savings from a reduced sample size are small. We estimate that the full experimental version of the research program will lead to a saving in research costs of $1,000,000 to $1,700,000 (Table 15.5).

Use of the experimental variant, however, requires that existing federal aid in the amounts of $15 million to $25 million (for the process study only) or $28 million to $50 million (for the full-scale project) must be diverted from present allocation procedures to the experimental design—a commitment that would have to be made by the Department of Health, Education, and Welfare. If this is done, the time schedule for the first year of the study becomes slightly tighter, and we would anticipate that the decision to commit federal aid funds to the problem will delay the entire project. We nevertheless feel that the experimental version is sufficiently superior to the standard design to make the management complications worthwhile.

Addition of a Panel of High School Graduates

We believe that the greatest weakness of the research program outlined in this report is the lack of specific recommendations for analysis of adult behavior resulting from racial experiences in childhood and adolescence. Although we believe this is the most important aspect of the controversy over the value of school desegregation, we chose not to make detailed recommendations in this area because existing data bases can be used more efficiently to meet these needs; new data do not have to be collected (see Sec. 14 above). These data bases should be investigated and contracts let for additional analysis. These contractors may recommend funding the collection of additional data. However, we cannot make recommendations until these data bases have been investigated in more detail and some analysis made. If, after that analysis, it is concluded that these data bases are not preferable to that which can be generated with this data base, the decision should be made to return to the 11th-12th grade panel of the large-scale longitudinal study and resurvey some of the students in that panel two or three years after graduation (in year 4 or 5 of the project). The original data collection will include detailed information to permit locating these students. We estimate the costs of a resurvey of high school alumni to be an additional $500,000 to $700,000, almost all of which will be data collection costs.

Study of Adult Racial Behavior

Another topic not adequately discussed in our report is the measurement of adult racial behavior as it has been affected by school racial experiences in earlier

holds that the main purpose of school desegregation is to reduce the prejudiced behavior of Anglo students when they become adults. Again, we have not made detailed recommendations because we do not know what data bases can best be used to deal with this problem. However, we strongly recommend investigating the possibility of attaching to parent surveys of the community reaction study a methodology for measuring the behavior of adult Anglo-Americans.

The methodological problems here are quite serious, but a solution would gather data of enormous value. In general, measurement of prejudice has been limited to measurement of attitudes, and much research has suggested that attitudes in and of themselves are poor cues to actual behavior (LaPierre 1934; Williams et al. 1959). The methodological difficulty lies in the fact that opportunities to either discriminate against minority groups or to combat discrimination are limited and not easily compared from one situation to another. One could, of course, create experimental stimuli, but this raises ethical problems and is again a poor substitute for observations for actual behavior (the hypothetical solution of a blue-collar worker to a problem faced by a personnel manager is not as interesting as the actual solution of the blue-collar worker to a racial problem that he has in fact had to deal with). We recommend that one of the client agencies pursue this problem, perhaps by letting a small contract for conceptualization of the problem and devising a research instrument, ideally one that can be combined with some ongoing survey such as the parent survey recommended in this study. It should be noted that the gathering of data on personal school experiences of public school teachers and principals in our survey (which is recommended in this study) does provide an excellent opportunity to link personal school racial experiences to adult behavior for one particular population.

Options to Permit Cost Savings

Undoubtedly many readers, bringing their own interests and hypotheses to bear on the research program outlined here, will recommend modifications and simplifications of this design.

One of the most obvious is that if the experimental design is not adopted, it would nevertheless be tempting to collapse the separate samples from the student panel and process substudies together. There are some difficulties in this. The problem is that each study ideally requires a different sample. The student panel survey should be a study of life in a typical sample of desegregated schools, with or without matched comparisons to typical segregated schools. On the other hand, the study of newly desegregated schools is just that—schools that are going through the initial stages of making the transition from segregation to desegregation. To assess the quality of education in desegregated schools from only these schools would be unfortunate.

The study of innovative elementary schools and high schools also requires a very special sample—that minority of schools which represent the most extreme versions of certain innovations. In this sample, secondary schools and elementary schools will not be sampled so that they serve the same attendance areas, and this is an important consideration of the student panel design. Nevertheless, it would be possible to modify the study of innovative schools and the student panel design so that the two could be based on an overlapping set of schools. However, the cost saving is not very large—only $300,000 to $500,000, and we are reluctant to recommend it.

The data collection contractor should be able to draw upon its considerable experience to recommend reductions in sample size and alternative methods of data

collection in order to reduce the costs. We are already recommending the use of mail and telephone questionnaires to parents, and self-administered questionnaires to both students and teachers in order to reduce costs.

Suppose that the various client groups collectively decide to sharply reduce the budget of this program?

The smallest possible project which would meet most of the goals of this study would consist of the student panel design only. A student panel, operating over four years in 122 schools at each grade level (64 segregated, 58 desegregated, in matched clusters) would provide good data on long-term school effects and fairly good data on the effects of school characteristics on student cognitive and noncognitive outcomes. It would provide some data on school innovations, some insight into the problems of newly desegregated schools, and good data on parental values. Data on Mexican-American and Puerto Rican students would vary in quality depending on sample size. Finally, data on community reaction to desegregation would be almost totally eliminated, although a small amount of funds for case studies might be helpful here. The study design is presented in Table 15.6.

This study, over four years, with a 30 percent sample of parents and a few case studies of desegregating districts analyzed along the lines described above, could be made for $4.5 to $7 million—about half the cost of the total program. Note that these estimated costs are reduced by only 50 percent, despite the fact that data collection has been reduced to about one-quarter of that proposed in the larger study. This reflects an effort to maintain as much of the analysis as possible, and it also reflects the loss of some important economies of scale in the data collection. Relative to the large-scale study, we would judge this design to perform as shown in the table below. The main problem is of course the sample size, which is adequate for measuring effects only of variables that can be measured in every school and that can be assumed to operate in the same manner across the entire sample. For example, this study would measure well the effects of staff racial attitudes on students, or the effects of racial composition. Studies that require locating relatively rare school attributes, such as unusual school innovations, would suffer most from the small sample size.

Task	*Performance*
Long-term school effects and desegregation effects on students	Very good
Description of school process factors in newly desegregated schools	Good
Location of promising innovations for desegregated schools	Poor
Evaluation of more common innovations	Fair
Evaluation of school effects on Mexican-American and Puerto Rican students	Fair to good

The study retains the quasi-experimental design sampling scheme, so that the chances of obtaining clear conclusions about the effects of desegregation on students remain high.

The elimination of the detailed study of community desegregation experience considerably weakens an important part of the study, although the existence of case study material and parent survey data permits some analysis of this area.

A variant of this study design which would permit it to make a strong contribution to policy regarding interventions and innovations in desegregated schools would include the experimental allocation of federally funded interventions in

Table 15.6

DESIGN FOR A LOWER-COST STUDY BASED ON STUDENT PANEL METHODOLOGY

Sampling: 29 matched quartets or quintets of schools at each grade level, drawn from matched sets of 40 school districts.[a]

Yields approximately 122 schools per grade level:

	22 black/Anglo northern schools
	6 black/Anglo/Puerto Rican schools
	6 black/Anglo/Mexican-American schools
	2 Puerto Rican/Anglo schools
	4 Mexican-American/Anglo schools
	18 southern black/Anglo schools
Subtotal	58 desegregated schools
	29 Anglo schools
	17 northern black schools
	4 Puerto Rican schools
	5 Mexican-American schools
	9 southern black schools
Subtotal	64 segregated schools

Data collection:

Spring year 1: students, teachers, principal, grades 3, 6, 9, +30 percent sample of all parents; use instruments designed for large-scale longitudinal study.

Spring year 2: students, teachers, principal, grades 4, 7, 10.

Spring year 3: students, teachers, principal, grades 2, 5, 8, 11.

Spring year 4: students, teachers, principal, grades 3, 6, 9, 12.

Each year, questionnaire is modified to gain better measures of school program: innovations, climate. Classroom observations are done each year.

Other data collection: case study, years 1-3, of 20 desegregated schools and their 10 districts.

Analysis: Main emphasis on long-term effects of desegregated school experience, heavy use of case study material to assist survey in explaining differential cognitive and non-cognitive growth.

Analysis of suburbanization of white parents in all 40 districts.

Analysis of school innovations is limited, includes synthesis of other data. Analysis of long-term change in districts studied by Coleman is possible.

---

[a]District sample is distributed as follows:

15 northern, stratified to include Puerto Rican in 3 or more
7 southern, studied by Coleman
8 other southern
5 with large Mexican-American populations.

School clusters are as follows:

North: 14 black/Anglo
3 Puerto Rican/Anglo
3 Mexican-American/Anglo
Note: 6 of above clusters will be tri-ethnic
9 southern black/Anglo.

desegregated schools during the second and third years of the project. This would require earmarking $7 million to $12 million in school aid, and would not alter the costs of the research. This study would essentially be a small version of the experimental process study described in Sec. 12, without a preceding large-scale study.

The presence of only 58 desegregated schools reduces the costs of the federal aid to schools; at the same time it limits the number of innovations that can be tested.

A second variant of this design is elimination of the segregated schools, thereby doubling the number of desegregated schools in the sample at the same cost. If both these variants are employed (the elimination of segregated schools and the use of the experiment) the result is to carry out the experimental process study of Sec. 12 without a preliminary large-scale longitudinal survey. This is another feasible variation on the basic design. In all cases, these variants use the basic building blocks of the main study—the theoretical base, the data collection method, and the questionnaire.

We recommend that a competent survey research firm be asked to make careful estimates of the costs of this project. We believe that such a firm could provide accurate estimates for each year of the project, which could greatly assist the Commission in choosing among the possibilities.

16. PROGRAM MANAGEMENT

In many ways, the single most difficult task in the entire program of research is the management of it. Research management is difficult in the best of circumstances, and this research will be especially hard to manage.

We have identified six management problems which will be difficult to solve:

1. Given that a large number of different contractors will be involved, how should their work be coordinated?
2. Given that a number of funding agencies may be involved, how should they be coordinated?
3. How can a highly satisfactory monitoring process be developed for a project as politically sensitive as this?
4. How can the project be made policy relevant?
5. How can the project be gotten off the ground?
6. How should the results of the project be translated into policy recommendations?

We will comment on each of these in turn.

COORDINATION OF CONTRACTORS

In Sec. 15 we identified 18 research tasks. These eighteen tasks can be combined into contracts in a large number of ways. In general, we recommend that the major analysis tasks from the data be given to different research teams on the grounds that any one research team tends to be limited by the amount of time available to senior researchers, and that requiring a single contractor to exercise a variety of different talents and interests unnecessarily limits the number of available researchers. In general, the strongest research teams can be obtained with contracts that are small and narrowly focused, that run for a long period of time, and that provide considerable lead time between circulation of the request for proposal and the actual beginning of work. This formula for contract awards will also make it possible to attract more qualified minority researchers, whose scarcity is a chronic problem in research on race relations.

Given such a large number of contractors, coordination becomes important. There will be a very large number of cases where one researcher's work is dependent on timely arrival of data or analysis from another contractor. We recommend three actions which will simplify problems of coordination.

1. All data should be gathered by a single contractor. Using one contractor gains considerable economy of scale, reduces the number of places where mistakes can be made, and increases accountability.

2. One contract, which includes assisting in the task of centralizing and coordinating the overall research program, should be let to a research group familiar with the problems of research and worthy of the respect of the other contractors. This management group (1) maintains the time schedule of the program—either makes sure that deadlines for various contractors are met or proposes alterations to the time schedule when deadlines slip; (2) serves as an intermediary between

clients and contractors, interceding in disputes, filling communication gaps, and breaking down communication barriers; (3) assists the client in the difficult task of translating policy needs into research problems; (4) lends technical expertise to the problem of choosing computer software and ensuring that technological problems in the multiple use of computer tapes are minimized.

3. The management team should hold a conference annually (or more often) to bring together researchers from the various research teams. This conference should operate in a very informal fashion with oral presentations and working papers from each research team providing the major method for exchanging hypotheses and sharing data problems. Good informal relationships between the members of the various research teams are critical. Contractors should be required to distribute informal working papers to the other research teams.

Other problems of coordination between contractors can be greatly minimized if contractors are chosen carefully. Because obtaining high-quality contractors is obviously important, we recommend that the usual very brief request for proposal time period be considerably lengthened. Highly skilled contractors are frequently not available on short notice to participate in a project. However, given the opportunity to schedule their work a number of months in advance, some of the most talented researchers in the country could be recruited for this project. The client will generally find that a moderate amount of freedom to alter the design of particular tasks can be granted to contractors without damaging the overall project.

Contractors who are selected for analysis tasks should have an early opportunity to interact with the data collection contractor, and that opportunity should be maintained throughout the life of data collection activities, since it is imperative that the analysis teams get to know the real quality of the data on a first-hand basis. They can only do this by acquiring an intimate understanding of the field-related real-life problems that the data collection contractor has faced.

COORDINATION OF FUNDING AGENCIES

Problems of coordination resulting from the involvement of more than one client agency need to be considered. This research program has elements within the research agenda of five agencies: the Office of Education, in the evaluation of federally funded programs for desegregated schools; the National Institute of Mental Health, because of its concern with the impact of desegregation on noncognitive and mental-health-related aspects of student behavior; the National Science Foundation, because of its concern with the impact of desegregation planning on community acceptance of desegregation; and, of course, the Commission on Civil Rights and the National Institute of Education have an interest in nearly all of the areas of the research program. Private foundations may also be interested in the effort.

We recommend that as many of these agencies as possible participate in the program, in order to eliminate duplication of services and to make maximum use of the data bank being created. If more than one agency does participate in the funding of the project, we recommend that a single client take responsibility for the data collection and the maintenance of the data bank while the others take charge of funding of analyses of the data.

The coordination of several funding agencies may require the creation of an interagency task force. This task force should consist of relatively senior decision-makers from each agency and should have a chairman with a great deal of experience. The task force should in turn report only to very senior policy-makers.

THE POLITICAL INSULATION OF THE PROGRAM

The creation of an interagency task force provides the opportunity to draw upon the staffs of a number of agencies for the project monitoring staff. The head project monitor should have at least five years experience in research managment, with special expertise in school desegregation, and familiarity with many research methodologies. He should be released from his own agency, so that he may report only to the chairman of the task force.

This monitoring staff has the ultimate responsibility of maintaining the scientific objectivity of this research and for this reason should be well insulated from ideological pressures and bureaucratic inferences within their agencies. The research monitor's function is to protect the research program from these pressures.

The contract let to a special management group discussed earlier is a device for providing staff assistance to the monitor.

MAINTAINING POLICY RELEVANCE

Although the research program must be insulated from political or ideological pressure, care must be taken to make certain that the project does not become merely academic. We suggest that a special subcommittee of policy-makers and researchers be appointed, reporting to the project monitor and the management group, whose sole function is to observe the work of all contractors identifying policy issues and policy instruments which grow out of all the research. This committee assists the monitor in deciding when to impose additional policy questions on each researcher.

INITIATING THE PROJECT

The project must be initiated in a situation of indeterminant funds and uncertain agency cooperation. Furthermore, the Commission on Civil Rights is the smallest of the various agencies that could be involved. In these circumstances, the seemingly simple task of making the initial decisions about the project becomes complex. One possible sequence of events is as follows:

1. The Commission first decides whether the general research area is worthy of pursuit at all.
2. The Commission contacts other agencies and recruit members for a temporary interagency advisory group. In the process of doing so, it should undertake the task of learning the history of other interagency large-scale research efforts in order to gain from their experience.
3. The Commission recruits, from its own or another agency's staff, a senior project monitor.
4. The senior project monitor works with the interagency group and the commission in developing more detailed cost estimates and isolates the initial set of policy decisions which must be made about the size, cost, and direction of the project.
5. The Commission and the interagency committee develop jointly a proposal for funding the effort.
6. The senior project monitor then recruits an advisory committee of senior researchers and agency staff persons. This advisory committee assists in locating

potential researchers and in developing requests for capability statements from them.

As funds appear, the monitor and his advisory committee (and the management group as soon as it is selected) begin the task of writing RFPs. The contract for data collection should be one of the first let, since the instrument included in Appendix C will need field testing and revision. Contracts for analysis should be let early enough to permit each analyst to participate in the final drafting of instruments relevant to his or her work.

The Interagency committee, the project monitor, and the research advisory committee will hopefully remain relatively unchanged throughout the life of the program.

A stronger set of RFPs can be written and a more competent set of contractors chosen if the capability statements are written based on an overview of the entire project, with each bidder indicating all of the portions of the project he or she is interested in. This permits the monitor to develop packages of tasks based on the known capability of bidders.

HOW SHOULD THE COMPLETED PROJECT BE TRANSLATED INTO POLICY RECOMMENDATIONS?

At each step in the project, the policy relevance of the work is maintained by the policy review committee mentioned earlier. This will not be sufficient in itself to make the project's final results relevant to national and local policy.

The research recommended in this report is certainly the most important on equality of educational opportunity since the Coleman study, and perhaps since the 1954 Supreme Court decision in *Brown* v. *The Topeka Board of Education.* But Mosteller and Moynihan (1972) argue that in the six years succeeding its publication the Coleman report had no effect on public policy with regard to equality of educational opportunity. It would be disastrous if the same ineffectiveness with regard to public policy were to characterize the research suggested in this report.

A review process should be begun as research reports begin to accummulate. This review process should include a wide variety of researchers policy-makers, and citizens and should be directed toward answering the following questions:

What is the weight of the evidence supporting various policy recommendations? Which ones can be unequivocally endorsed and which need further analysis?

Where contradictions appear between different research reports, what is the source of the contradiction, is the issue important enough to merit an effort at resolution, and how can it be resolved?

Where criticisms have been made of the research, are the criticisms merited, and what needs to be done?

This process feeds back to the research, in many cases while the research is still under way, to try to bring about scientific consensus on the results. It will of course not succeed in doing so, but it should help.

As the program comes to a close, the policy process turns to producing an overall report to the nation on what we have learned, and this new knowledge affects policy.

BIBLIOGRAPHY

Agger, Robert E., and Marshall N. Goldstein, *Who Will Rule the Schools: A Cultural Class Crisis,* Wadsworth, Belmont, California, 1971.

Alford, Robert R., "The Comparative Study of Urban Politics," *The Search for Community Power,* Willis D. Hawley and Frederick M. Wirt (eds.), Prentice-Hall, Englewood Cliffs, New Jersey, 1968.

Alford, Robert R., and Eugene C. Lee, "Voting Turnout in American Cities," *American Political Science Review,* Vol. 62, pp. 796-813, September 1968.

Allport, Gordon W., *The Nature of Prejudice,* Addison-Wesley, Reading, Massachusetts, 1954.

Almy, Timothy, and Harlan Hahn, "Perceptions of Educational Conflict," *Education and Urban Society,* Vol. 3, pp. 440-452, August 1971.

Amidon, Edmund J., and Ned A. Flanders, *The Role of the Teacher in the Classroom,* pp. 1-101, Association for Productive Teaching, Minneapolis, Minnesota, 1971.

Amir, Yehuda, "Contact Hypothesis in Ethnic Relations," *Psychological Bulletin,* Vol. 71, pp. 319-342, No. 5, May 1969.

Anderson, Gary J., *The Assessment of Learning Environments: A Manual for the Learning Environment Inventory and the My Class Inventory,* Atlantic Institute of Education, Halifax, Nova Scotia, Canada, September 1973.

Armor, David J., "The Evidence of Busing," *The Public Interest,* Vol. 28, pp. 90-126, Summer 1972.

——, "The Double Double Standard: A Reply," *The Public Interest,* Vol. 30, pp. 119-131, Winter 1973.

Bachman, Gerald G., *Young Men in High School and Beyond: A Summary of Findings from the Youth in Transition Project,* Final Report for Contract OE-5-85-054, U.S. Office of Education, 1972.

Bachman, Gerald G., et al., *Youth in Transition,* Institute for Social Research, University of Michigan, Ann Arbor, 1971.

"Backlash in Boston—And Across the U.S.," *Newsweek,* November 6, 1967.

Bagwell, William, *School Desegregation in the Carolinas: Two Case Studies,* University of South Carolina Press, Columbia, 1972.

Banfield, Edward C., and James Q. Wilson, *City Politics,* Vintage Books, New York, 1963.

Bendiner, Robert, *The Politics of Schools,* Harper and Row, New York, 1969.

Bettelheim, Bruno, and Morris Janowitz, *Social Change and Prejudice,* The Free Press, New York, 1964.

——, "Middle-Class Teacher and Lower-Class Child," *Rethinking Urban Education,* Herbert J. Walberg and Andrew T. Kopan (eds.), Jossey-Bass, San Francisco, pp. 3-11, 1973.

Blalock, Hubert M., *Causal Inferences in Nonexperimental Research,* The University of North Carolina Press, Chapel Hill, 1961.

——, *Social Statistics,* McGraw-Hill, New York, 1960.

——, *Toward a Theory of Minority-Group Relations,* John Wiley and Sons, New York, 1967.

Blau, Peter M., and W. Richard Scott, *Formal Organization,* Chandler, San Francisco, 1962.

Bloom, Benjamin S., et al., *Compensatory Education for Cultural Deprivation,* Holt, Rinehart and Winston, New York, 1965.

Bogardus, Emory S., "Racial Distance Changes in the United States During the Past 30 Years," *Sociology and Social Research,* Vol. 43, pp. 127-134, November 1958.

Bohrnstedt, George W., and T. Michael Carter, "Robustness in Regression Analysis," Herbert L. Costner (ed.), *Sociological Methodology,* Jossey-Bass, San Francisco, pp. 118-146, 1971.

Bonacheck, Edna, and Robert F. Goodman, *Deadlock in School Desegregation,* Praeger, New York, 1972.

Boocock, Lorane S., "Toward a Sociology of Learning: A Selective Review of Existing Research," *Sociology of Education,* Vol. 39, pp. 1-45, 1966.

Boudon, Raymond, "Propriétés Individuelles et Propriétés Collectives: Un Problème d'Analyse Ecologique," *Review Francaise de Sociologie,* Vol. 4, pp. 275-299, July-September 1963.

Bouma, Donald A., and James Hoffman, *The Dynamics of School Integration,* Eerdmans, Grand Rapids, Michigan, 1968.

Boyd, Richard, "Popular Control of Public Policy: A Normal Vote Analysis of the 1968 Election," *American Political Science Review,* Vol. 66, pp. 429-449, June 1972.

Boyd, William L., "Community Status and Conflict in Surburban School Politics," paper presented at the American Educational Research Association Annual Meeting, Chicago, April 3-7, 1972.

Brewer, Garry, *Politicians, Bureaucrats, and the Consultant,* Basic Books, New York, 1973.

Brophy, Jere E., and Thomas L. Good, *Teacher Child Dyadic Interaction: A Manual for Coding Classroom Behavior,* Report Series 27, The Research and Development Center for Teacher Education, The University of Texas, Austin, December 1969.

——, *Teachers' Communication of Differential Expectations for Children's Classroom Performance: Some Behavioral Data,* Report Series 25, The Research and Development Center for Teacher Education, The University of Texas, Austin, 1969.

Bryant, Edward C., H. O. Hartley, and R. J. Jessen, "Design and Estimation in Two-Way Stratification," *Journal of the American Statistical Association,* Vol. 55, No. 289, pp. 105-124, March 1960.

Bureau of the Census, *Statistical Abstract of the United States, 1972,* U.S. Government Printing Office, Washington, D.C., 1972.

Burnham, Walter Dean, *Critical Elections and the Mainsprings of American Politics,* W. W. Norton, New York, 1970.

Cain, Glen G., and Harold W. Watts, "Problems in Making Policy Inferences from the Coleman Report," *American Sociological Review,* Vol. 35, pp. 228-242, April 1970.

California State Department of Education, Bureau of Intergroup Relations, *Racial and Ethnic Survey of California Public Schools,* 1969.

Campbell, Angus, *White Attitudes Toward Black People,* Institute for Social Research, Ann Arbor, 1971.

Campbell, Angus, et al., *The American Voter: An Abridgement,* John Wiley and Sons, New York, 1964.

Campbell, Angus, and Howard Schuman, "Racial Attitudes in Fifteen American Cities," *Supplemental Studies for the National Advisory Commission on Civil Disorders,* U.S. Government Printing Office, Washington, D.C., 1968.

Carter, Richard F., and William G. Savard, *Influence of Voter Turnout on School Bond and Tax Turnout Elections,* U.S. Government Printing Office, Washington, D.C., 1961.

Carver, Frederick D., "Social Class and School Board Role Expectations," *Urban Education,* Vol. 3, pp. 143-154, 1968.

Chandler, Marjorie O., "The Longitudinal Study of Educational Effects: Design of the Follow-up Surveys," paper presented at the American Association for the Advancement of Science, San Francisco, March 1974.

Chesler, Mark A., and Phyllis Segal, *Characteristics of Negro Students Attending Previously All-White Schools in the Deep South,* Center for Research on Utilization of Scientific Knowledge, Institute for Social Research, University of Michigan, Ann Arbor, September 1967.

Chesler, Mark A., "Interaction and Teaching Alternatives in Desegregated Classrooms," *Racial Crisis in American Education,* Robert L. Green (ed.), Follett Educational Corporation, Chicago, 1969.

Chesler, Mark A., C. Jorgensen, and P. Erenberg, *Planning Educational Change: Integrating the Desegregated School,* Vol. III, Office of Education, U.S. Government Printing Office, Washington, D.C., 1970.

Clark, Carl A., "Making School Rewarding," *Rethinking Urban Education,* Herbert J. Walberg and Andrew T. Kopan (eds.), Jossey-Bass, San Francisco, 1972.

Clark, Terry N., *Comparative Community Politics,* Sage, Los Angeles, 1973.

——, "Community Structure and Decisionmaking," *Community Structure and Decisionmaking: Comparative Analyses,* Chandler, San Francisco, 1968.

——, "Community Structure, Decisionmaking, Budget Expenditures, and Urban Renewal in 51 American Communities," *American Sociological Review,* Vol. 33, pp. 576-593, August 1968.

Clark, James, "Environment, Process, and Policy: A Reconsideration," *American Political Science Review,* Vol. 63, pp. 1172-1182, December 1969.

Cohen, A. D., "Innovative Education for La Raza: A Sociolinguistic Assessment of a Bilingual Education Program in California," Ph.D. Dissertation, Stanford University, December 1972.

Cohen, Elizabeth G., "Sociology and the Classroom: Setting the Conditions for Teacher-Student Interactions," *Review of Educational Research,* Vol. 42, No. 4, pp. 441-452.

Coleman, James S., *Community Conflict,* The Free Press, New York, 1957.

Coleman, et al., *Equality of Educational Opportunity,* U.S. Government Printing Office, Washington, D.C., 1966.

Coleman, James S., "Social and Cultural Integration and Educational Policy," *Rethinking Urban Education,* Herbert J. Walberg and Andrew T. Kopan (eds.), Jossey-Bass, San Francisco, 1972.

Coles, Robert, *Children of Crisis: A Study of Courage and Fear,* Dell, New York, 1967.

Counts, George S., *The Social Composition of Boards of Education,* The University of Chicago Press, Chicago, 1927.

Crain, Robert L., "School Integration and the Academic Achievement of Negroes," *American Journal of Sociology,* Vol. 75, pp. 593-606, 1970.

Crain, Robert L., and Carol Sachs Weisman, *Discrimination, Personality, and Achievement,* Seminar Press, New York, 1972.

Crain, Robert L., et. al., "Report to the Office of Education on the Causes and Effects of School Board Recruitment Patterns," The Johns Hopkins University, Baltimore, Maryland, 1972 (mimeographed).

Crain, Robert L., et al., *The Politics of School Desegregation,* Aldine, New York, 1968.

Crain, Robert L., Elihu Katz, and Donald B. Rosenthal, *The Politics of Community Conflict: The Fluoridation Decision,* Bobbs-Merrill, .Indianapolis, Indiana, 1969.

Crain, Robert L., Laura Morlock, and James J. Vanecko, "The Influence of Reputational, Positional and Decisional Elites in Northern Cities," paper presented at the American Sociological Association Meeting, San Francisco, September 1969.

Crain, Robert L., and Donald B. Rosenthal, "Structure and Values in Local Political Systems: The Case of Fluoridation Decisions," *Journal of Politics,* Vol. 8, pp. 169-195, February 1966.

Cronin, Joseph, *The Control of Urban Schools,* The Free Press, New York, 1973.

Cunningham, Luverne L., "Decision-Making Behavior of School Boards," *American School Board Journal,* Vol. 164, pp. 13-15, February 1962.

Dahl, Robert, "A Critique of the Ruling Elite Model," *The American Political Science Review,* Vol. 52, pp. 463-469, June 1958.

Danzger, Herbert M., "On the Use of Newspaper Reports as a Source of Data," *Social Science Methods,* Robert B. Smith (ed.), The Free Press, New York, 1974.

——, "Critical Mass and Historical Process as Factors in Community Conflict," paper presented at the annual meeting of the American Sociological Association, Washington, D.C., September 1970.

——, "A Quantified Description of Community Conflict," *American Behavioral Scientist,* Vol. 12, No. 2, pp. 9-15, November-December 1968.

——, "Civil Rights Conflict and Community Power Structure," Ph.D. Dissertation, Columbia University, 1968.

Davis, James A., et al., "A Technique for Analyzing the Effects of Group Composition," *American Sociological Review,* Vol. 26, pp. 215-225, April 1961.

——, "Busing," Working Paper No. 5 in *Southern Schools: An Evaluation of the Effects of the Emergency School Assistance Program and School Desegregation,* Vol. 2, National Opinion Research Center, Chicago, 1973.

Dawson, Richard E., and James A. Robinson, "Interparty Competition, Economic Variables, and Welfare Politics in the American States," *Journal of Politics,* Vol. 25, pp. 265-289, May 1963.

Deutsch, Martin, "Social Intervention and the Malleability of the Child," *The Disadvantaged Child,* Martin Deutsch (ed.), Basic Books, New York, pp. 7-9, 1967.

Dolbeare, Kenneth, *The School Prayer Decisions: From Court Policy to Local Practice,* The University of Chicago Press, Chicago, 1971.

——, "The Supreme Court and the States: From Abstract Doctrine to Local Behavioral Conformity," *The Impact of Supreme Court Decisions,* Theodore L. Becker (ed.), Oxford University Press, New York, pp. 206-213, 1969.

Dolbeare, Kenneth, and Phillip E. Hammond, "The Political Party Basis of Attitudes Toward the Supreme Court," *Public Opinion Quarterly,* Vol. 32, pp. 16-30, Spring 1968.

Downs, Anthony, *Urban Problems and Prospects,* Dorsey Press, Markham, Illinois, 1970.

Dreeben, Robert, *On What is Learned in School,* Addison-Wesley, Reading, Massachusetts, 1968.

Duncan, Otis Dudley, Ray P. Cuzzort, and Beverly Duncan, *Statistical Geography: Problems in Analyzing Areal Data,* The Free Press, Glencoe, Illinois, 1961.

Dye, Thomas, *Politics, Economics, and the Public,* Rand McNally, Chicago, 1966.

——, "Urban School Segregation: A Comparative Analysis," *Urban Affairs Quarterly,* Vol. 4, pp. 141-165, December 1968.

Eash, Maurice F., and James T. Napolitan, "Insights and Alternatives from a Black Street Academy," *Rethinking Urban Education,* Herbert J. Walberg and Andrew T. Kopan (eds.), Jossey-Bass, San Francisco, 1972.

Ehrlich, Howard J., "The Social Psychology of Reputations for Community Leadership," *The Sociological Quarterly,* Vol. 8, pp. 514-530, Summer 1967.

——, *The Social Psychology of Prejudice,* John Wiley and Sons, New York, 1972.

Eitzen, D. Stanley, "Social Class, Status Inconsistency and Political Attitudes," *Political Attitudes and Public Opinion,* Dan D. Nimmo and Charles M. Bonjean (eds.), David McKay, New York, 1972.

Elder, Glen H., Jr., *Adolescent Socialization and Personality Development,* Rand McNally, New York, 1971.

Emmer, Edmund T., *Classroom Observation Scales,* The Research and Development Center for Teacher Education, The University of Texas, Austin, 1971.

Emrick, John A., et al., *Interim Evaluation of the National Follow Through Program 1969-1971,* Stanford Research Institute, Report SRI-URU-7370, February 1973.

Engle, P. L., "The Use of the Vernacular Languages in Education: Revisited," an unpublished literature review prepared for the Ford Foundation Office of Mexico, Central America, and the Caribbean, University of Illinois at Chicago Circle, 1973.

Eulau, Heinz, and Robert Eyestone, "Policy Maps of City Councils and Policy Outcomes: A Developmental Approach," *American Political Science Review,* Vol. 60, pp. 640-654, September 1966.

Fiedler, Fred, *Leader Attitudes and Group Effectiveness,* University of Illinois Press, Urbana, 1973.

Fischer, John H., "The School Park," *Education Parks,* Clearinghouse Publications, No. 9, U.S. Government Printing Office, Washington, D.C., October 1967.

Fisher, Ronald A., *Design of Experiments,* 4th ed., Oliver and Boyd, Edinburgh, 1935.

Fishman, Joshua A., and Robert L. Cooper, "Alternative Measures of Bilingualism," *Journal of Verbal Learning and Verbal Behavior,* Vol. 8, No. 21, pp. 276-282, April 1969.

Fishman, Joshua A., Robert L. Cooper, and R. Ma, *Bilingualism in the Barrio,* Indiana University Publications Language Science Monographs, Vol. 7, Bloomington, 1971.

Flanagan, John C., *Five Years After High School,* American Institutes for Research, Final Report, Grant OEG-0-9-610065-1367(085), U.S. Office of Education, 1971.

Flanders, Ned A., *Analyzing Teaching Behavior,* Addison-Wesley, Reading, Massachusetts, 1970.

——, *Teacher Influence, Pupil Attitudes and Achievement,* HEW-OE-25040, U.S. Government Printing Office, Washington, D.C., 1965.

Flannery, J. Harold, "Busing and Metropolitan Desegregation," *Developments in School Desegregation and the Law,* Charles Moody, Charles Vergon, and John Taylor (eds.), University of Michigan School of Education, Ann Arbor, 1972.

——, "Unconstitutional Segregation and Court Ordered Desegregation," *Developments in School Desegregation and the Law,* Charles Moody, Charles Vergon, and John Taylor (eds.), University of Michigan School of Education, Ann Arbor, 1972.

Foster, Gordon, "Desegregating Urban Schools: A Review of Techniques," *Harvard Educational Review,* Vol. 43, No. 1, pp. 5-36, February 1973.

Frankel, Lester R., and J. Stevens Stock, "On the Sample Survey of Unemployment," *Journal of the American Statistical Association,* Vol. 37, pp. 77-80, 1942.

Friedman, Daniel J., "Happiness is a Warm Gun: A Theoretical and Empirical Perspective on White Militancy in Boston," Ph.D. Dissertation, University of North Carolina, 1972.

——, *White Militancy in Boston: A Reconsideration of Marx and Weber,* D. C. Heath, Lexington, Massachusetts, 1973.

Fuchs, Estelle, *Pickets at the Gates,* The Free Press, New York, 1966.
Fuller, Frances F., *FAIR System Manual: Fuller Affective Interaction Records,* The Research and Development Center for Teacher Education, The University of Texas, Austin, 1969.
Gage, Nathaniel L. (ed.), *Handbook of Research on Teaching,* Rand McNally, Chicago, 1963.
Gamson, William A., "Rancorous Conflict in Community Politics," *American Sociological Review,* Vol. 31, February 1966.
Gardner, R. C., and Wallace E. Lambert, *Attitudes and Motivation in Second Language Learning,* Newbury House, Rowley, Massachusetts, 1972.
Gates, Robbins L., *The Making of Massive Resistance: Virginia's Politics of Public School Desegregation, 1954-1956,* University of North Carolina Press, Chapel Hill, 1964.
Gerard, Harold B., and Norman Miller, "Factors Contributing to Adjustment and Achievement in Racially Desegregated Public Schools," University of California, Los Angeles, 1971 (mimeograph).
Gilbert, Claire W., "Community Power and Decisionmaking: A Quantitative Examination of Previous Research," *Community Structure and Decisionmaking: Comparative Analyses,* Terry N. Clark (ed.), Chandler, San Francisco, 1968.
Glassberg, Andrew, "The Linkage Between Urban Policy Outputs and Voting Behavior: New York and London," paper presented at the American Political Science Association Meeting, Washington, D.C., September 1972.
Gold, David, "Independent Causation in Multivariate Analysis: The Case of Political Alienation and Attitudes Toward a School Bond Issue," *American Sociological Review,* Vol. 27, pp. 85-87, February 1962.
Goldhammer, Keith, *The School Board,* The Center for Applied Research in Education, Inc., New York, 1964.
Good, Thomas L., and Jere E. Brophy, "Teacher-Child Dyadic Interactions: A New Method of Classroom Observation," *Journal of School Psychology,* Vol. 8, No. 2, pp. 131-138, 1970.
——, "Analyzing Classroom Interaction: A More Powerful Alternative," Report No. 26, The Research and Development Center for Teacher Education, University of Texas, Austin, 1969.
Goodlad, John I., *Desegregating the Integrated School,* Clearinghouse Publications, No. 9, U.S. Government Printing Office, Washington, D.C., October 1967.
Goodlad, John I., and Robert H. Anderson, *The Nongraded Elementary School,* Harcourt Brace and World, New York, 1963.
Goodman, Leo A., "Some Alternatives to Ecological Correlation," *American Journal of Sociology,* Vol. 64, pp. 610-625, May 1959.
Goodman, Rol, and Leslie Kish, "Controlled Selection—A Technique in Probability Sampling," *Journal of American Statistical Association,* Vol. 45, No. 251, pp. 350-372, September 1950.
Goodman, Rol, "A Controlled Selection Which Permits Unbiased Estimates of Sampling Variances," Bureau of the Census, Washington, D.C., September 10, 1963 (mimeograph).
Gordon, Ira, *Early Child Stimulation Through Parent Education,* final report to the Children's Bureau for Development of Human Resources, University of Florida, College of Education, Gainesville, 1969.
Graham, Hugh D., *Crisis in Print: Desegregation and the Press in Tennessee,* Vanderbilt University Press, Nashville, Tennessee, 1967.
Greeley, Andrew M., "Political Attitudes Among American White Ethnics," *Public Opinion Quarterly,* Vol. 36, pp. 213-220, Summer 1972.

——, *That Most Distressful Nation: The Taming of the American Irish,* Quadrangle Books, Chicago, 1972.

Greeley, Andrew M., and Paul Sheatsley, "Attitudes Toward Racial Integration," *Scientific American,* Vol. 225, pp. 10, 13-19, 120, December 1971.

Green, Donald R., and J. F. Draper, "Exploratory Studies of Bias in Achievement Tests," paper presented at the Annual Meeting of the American Psychological Association, September 1972.

Green, James A., and Harold B. Gerard, "School Desegregation and Ethnic Attitudes," *Integrating the Organization,* H. Fromkin and J. Sherwood (eds.), The Free Press, Glencoe, Illinois.

Gross, Neal, Joseph Giaqwinta, and Marilyn Bernstein, *Implementing Organizational Innovation: A Sociological Analysis of Planned Educational Change,* Basic Books, New York, 1971.

Hadden, Jeffrey, and Edward Borgatta, *American Cities: Their Social Characteristics,* Rand McNally, Chicago, 1965.

Hahn, Harlan, "Correlates of Public Sentiments about War: Local Referenda on the Vietnam Issue," *American Political Science Review,* Vol. 64, pp. 1186-1198, December 1970.

——, "Ecological Data and Structural Characteristics: Some Notes on the Homogeneity of Precincts and Census Tracts," paper presented at the Western Political Science Association Convention, Albuquerque, New Mexico, April 1971.

——, "Ethos and Social Class: Referenda in Canadian Cities," *Polity,* pp. 295-315, Spring 1970.

——, "Ethnic Politics and Racial Issues," *Western Political Quarterly,* Vol. 24, pp. 719-730, December 1971.

——, "Northern Referenda on Fair Housing: The Response of White Voters," *Western Political Quarterly,* Vol. 21, pp. 483-495, September 1968.

——, "Socioeconomic Characteristics and Electoral States: A Comparative Study of Community Referendums," paper presented at the Western Political Science Association Meeting, Sacramento, California, April 1970.

Haller, A. O., and I. W. Miller, *The Occupational Aspiration Scale: Theory, Structure, and Correlates,* 2nd ed., Schenkman, Cambridge, Massachusetts, 1971.

Haller, A. O., et al., "Level of Occupational Aspiration: An Empirical Analysis, *American Sociological Review,* Vol. 39, pp. 113-121, 1974.

Hamilton, Howard D., "Voting Behavior in Open Housing Referenda," *Political Attitudes and Public Opinion,* Dan Nimmo and Charles M. Bonjean (eds.), David McKay, New York, 1972.

Hamilton, Richard F., "Black Demands, White Reactions, and Liberal Alarms," *Blue Collar Workers: A Symposium on Middle America,* Sar A. Levitan (ed.), McGraw-Hill, New York, 1971.

——, *Class and Politics in the United States,* John Wiley and Sons, New York, 1972.

——, "Class and Race in the United States," *The Revival of American Socialism,* George Fischer (ed.), Oxford University Press, New York, 1971.

——, "Liberal Intelligentsia and White Backlash," *Dissent,* pp. 225-232, Winter 1972.

Handlin, Oscar, "The Goals of Integration," *Daedalus,* Vol. 95, No. 1, pp. 268-286, Winter 1966.

Hannan, Michael T., *Problems of Aggregation and Disaggregation in Sociological Research,* University of North Carolina, Institute for Research in Social Science, D. C. Heath, Chapel Hill, 1972.

Hansen, Morris H., and W. N. Hurwitz, "On the Theory of Sampling from Finite Populations," *Annals of Mathematical Statistics,* Vol. 14, pp. 333-362, 1943.

Hansen, Morris H., W. N. Hurwitz, and W. G. Madow, *Sample Survey Methods and Theory: Methods and Applications,* John Wiley and Sons, New York, 1953.

Hanushek, Eric, "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data, *American Economic Review,* Vol. 61, No. 2, pp. 280-288, May 1971.

Harrison, Forest, and Conrad Briner, *Implications of the Organizational Climate Description Questionnaire for Research and Practice,* Claremont Graduate School, October 1972.

Harvard-MIT Joint Center, *How People View Their City,* The Joint Center, Cambridge, Massachusetts, 1970.

Hatley, Richard V., and Martin Burlingame, "Voting Behavior in Four Albuquerque School Financial Referenda," *Education and Society,* Vol. 4, pp. 293-311, May 1970.

Hauser, Robert Mason, *Socioeconomic Background and Educational Performance, Washington,* American Sociological Association, 1972.

Hawkes, Thomas H., and Norma F. Furst, "Research Note: Race, Socioeconomic Status, Achievement, I.Q., and Teachers Rating of Behavior as Factors Relating to Anxiety in Upper Elementary School Children," *Sociology of Education,* Vol. 15, pp. 333-350, 1971.

Hawley, Amos H., "Community Power and Urban Renewal Success," *American Journal of Sociology,* Vol. 68, pp. 422-431, January 1963.

Hess, Irene, and K. S. Srikantan, "Some Aspects of the Probability Sampling Technique of Controlled Selection," *Health Services Research,* Vol. 1, No. 1, pp. 8-52, Summer 1966.

Hess, Robert, et al., *The Cognitive Environments of Urban Preschool Children: The Follow-Up Phase,* The Graduate School of Education, The University of Chicago, 1969.

Hill, Roscoe, and Malcolm Feeley (eds.), *Affirmative School Integration: Efforts to Overcome De Facto Segregation in Urban Schools,* Sage, Beverly Hills, California, 1968.

Hillary, George A., "Definitions of Community: Areas of Agreement," *Rural Sociology,* Vol. 20, pp. 194-204, 1955.

Hilton, Thomas L., et al., *Stability and Instability in Academic Growth,* Educational Testing Service Final Report for Grant OEG-2-70013(509), U.S. Office of Education, 1971.

Hilton, Thomas L., and C. Patrick, "Cross-sectional Versus Longitudinal Data: An Empirical Comparison of Mean Differences in Academic Growth," *Journal of Educational Measurements,* Vol. 7, pp. 15-24, Spring 1970.

Hirsch, Herbert, and Donohew, Lewis, "A Note on Negro-White Differences in Attitudes Toward the Supreme Court," *Social Science Quarterly,* Vol. 49, pp. 557-562, December 1968.

Hoehn, Arthur J., "A Study of Social Status Differentiation in the Classroom Behavior of Nineteen Third Grade Teachers," *The Journal of Psychology,* Vol. 39, pp. 269-292, 1954.

Hofferbert, Richard, "The Relation Between Public Policy and Some Structural and Environmental Variables in the American States," *American Political Science Review,* Vol. 60, pp. 73-82, March 1966.

Holden, Ann, *The Bus Stops Here,* Agathon Press, New York, 1974.

Horton, John E., and Wayne E. Thompson, "Powerlessness and Political Negativism: A Study of Defeated Local Referendums," *American Journal of Sociology,* Vol. 67, pp. 485-493, March 1962.

Hyman, Herbert H., and Paul Sheatsley, "Attitudes Toward Desegregation," *Scientific American,* Vol. 211, pp. 16-23, July 1964.

Irwin, Galen A., and Duane A. Meeter, "Building Voter Transition Models for Aggregate Data," *Midwest Journal of Political Science,* Vol. 13, pp. 545-566, November 1969.

Jackman, Mary R., and Robert W. Jackman, "An Interpretation of the Relation Between Objective and Subjective Social Status," *American Sociological Review,* October 1973.

James, Thomas H., James A. Kelly, and Walter Garms, "The School Budget Process in Large Cities," *The Politics of Education at the Local, State and Federal Level,* Michael W. Kirst (ed.), McCutcheon, Berkeley, 1970.

Jencks, Christopher S., "The Coleman Report and the Conventional Wisdom," *On Equality of Educational Opportunity,* Frederick Mosteller and Daniel P. Moynihan (eds.), Vintage Books, New York, pp. 69-115, 1972.

——, *Inequality,* Basic Books, New York, June 1972.

Jennings, M. Kent, and Harmon Zeigler, "Response Styles and Politics: The Case of School Boards," *Midwest Journal of Political Science,* Vol. 15, pp. 290-321, May 1971.

Jessen, R. J., "Some Properties of Probability Lattice Sampling," *Journal of the American Statistical Association,* Vol. 68, pp. 20-28, 1973.

——, "Probability Sampling with Marginal Restraints," *Journal of the American Statistical Association,* Vol. 65, No. 330, pp. 776-796, June 1970.

Katz, Daniel, and Robert L. Kahn, *The Social Psychology of Organization,* John Wiley and Sons, New York, 1966.

Katz, Irwin, "Factors Influencing Negro Performance in the Desegregated School," *Social Class, Race and Psychological Development,* Martin Deutsch, Irwin Katz, Arthur R. Jensen (eds.), Holt, Rinehart and Winston, New York, pp. 254-292, 1968.

——, "Desegregation or Integration in Public Schools? The Policy Implications of Research," *Integrated Education,* Issue 30, Vol. 5, No. 6, pp. 15-27, December 1967-January 1968.

Katzman, Martin T., *The Political Economy of Urban Schools,* Harvard University Press, Cambridge, Massachusetts, 1971.

Kaufman, Harold F., "Toward an Interactional Conception of Community," *Social Forces,* Vol. 38, pp. 8-17, October 1959.

Kelley, Jonathan, "Racism and School Busing," to be published in *Public Opinion Quarterly,* Columbia University, 1973 (lithographed).

Kerlinger, Fred, *Review of Research in Education,* Vol. 1, F. E. Peacock, Itasca, Illinois, 1973.

Kessel, John H., "Public Perceptions of the Supreme Court," *Midwest Journal of Political Science,* Vol. 10, pp. 167-191, May 1966.

King, Bert, and Elliot McGinnis, *Attitudes, Conflict and Social Change,* Academic Press, New York, 1972.

King, Gary W., and Walter E. Freeman, *Conflict Over Schools,* The Institute for Community Development, Michigan State University, Lansing, 1963.

Kirby, David J., T. Robert Harris, and Robert L. Crain, *Political Strategies in Northern School Desegregation,* D. C. Heath, New York, 1973.

Kirst, Michael W. (ed.), *The Politics of Education at the Local,* State and National Level, McCutcheon, Berkeley, California, 1970.

Koht, Herbert R., *The Open Classroom,* A New York Review Book, New York, 1969.

Kozol, Jonathan, *Death at an Early Age: The Destruction of the Hearts and Minds*

of Negro Children in the Boston Public Schools, Houghton-Mifflin, Boston, 1967.

Kranz, P. L., Wilford A. Weber, and Kenneth N. Fishell, "The Relationships Between Teacher Perception of Pupils and Teacher Behavior Toward Those Pupils," paper presented at the annual meeting of the American Educational Research Association, Minneapolis, March 1970.

L'Abato, Luciano, Yvonne Oslin, and Vernon W. Stone, "Educational Achievement," *Comparative Studies of Blacks and Whites in the United States,* Kent S. Miller and Ralph Mason Druger (eds.), Seminar Press, New York, pp. 325-354, 1967.

Labovitz, Sanford, "Criteria for Selecting a Significance Level: A Note on the Sacredness of .05," *Stages of Social Research,* Dennis P. Forcese and Stephen Richter (eds.), Prentice Hall, Englewood Cliffs, New Jersey, 1970.

LaBrecque, Richard T., "Policy Implications of the Coleman Report Reanalysis," *Phi Delta Kappan,* pp. 606-609, May 1973.

LaNoue, George R., and Bruce R. Smith, *The Politics of School Decentralization,* D. C. Heath, New York, 1973.

Ladefoged, Peter, R. Glick, and C. Criper, *Language in Uganda,* Oxford University Press, Nairobi, 1972.

Lambert, Wallace E., and G. R. Tucker, *Bilingual Education of Children: The St. Lambert Experience,* Newbury House, Rowley, Massachusetts, 1972.

Leibow, Elliot T., *Tally's Corner: A Study of Negro Streetcorner Men,* Little, Brown and Company, Boston, 1967.

Lemon, Richard, *The Troubled American,* Simon and Schuster, New York, 1970.

Levy, Frank, *Northern Schools and Civil Rights: The Racial Imbalance Act of Massachusetts,* Markham, Chicago, 1971.

Lewis, Anthony, *Portrait of a Decade,* Random House, New York, 1964.

——, *Race Relations in the U.S., 1954-1968,* Charles Scribner and Sons, New York, 1970.

Lewis, Oscar, *La Vida: A Puerto Rican Family in the Culture of Poverty—San Juan and New York,* Random House, New York, 1966.

Likert, Rensis, *The Human Organization,* McGraw-Hill, New York, 1967.

Lineberry, Robert L., "Approaches to the Study of Community Politics," *Community Politics,* Charles Bonjean, Terry N. Clark, and Robert L. Lineberry (eds.), The Free Press, New York, 1971.

Lineberry, Robert L., and Edmund P. Fowler, "Reformism and Public Policies in American Cities," *The American Political Science Review,* Vol. 61, pp. 701-716, September 1967.

Lineberry, Robert L., and Ira Sharkansky, *Urban Politics and Public Policy,* Harper and Row, New York, 1971.

Lipsky, Michael, "Protest as a Political Resource," *The American Political Science Review,* Vol. 62, pp. 1144-1158, December 1968.

Lopreato, Joseph, and Lawrence E. Hazelrigg, *Class, Conflict and Mobility,* Chandler, San Francisco, 1972.

Lopreato, Joseph, and Janet S. Chafetz, "The Political Orientation of Skidders: A Middle Range Theory," *American Sociological Review,* Vol. 35, pp. 440-450, June 1970.

Lord, Frederic M., and M. R. Novick, *Statistical Theories of Mental Test Scores,* Addison-Wesley, Reading, Massachusetts, 1968.

Lyons, Robert, "Racial Imbalance and State Desegregation Standards," *Developments in School Desegregation and the Law,* Charles Moody, Charles Vergon,

and John Taylor (eds.), University of Michigan School of Education, Ann Arbor, 1972.

Mack, Raymond W. (ed.), *Our Children's Burden: Studies of Desegregation in Nine American Communities,* Random House, New York, 1968.

Mann, F. C., "Toward an Understanding of the Leadership Role in Formal Organizations," *Leadership and Productivity,* Robert Dubin, G. Homans, and D. Miller (eds.), Chandler, San Francisco, 1965.

Martin, Roscoe, *Government and the Suburban School,* Syracuse University Press, Syracuse, New York, 1962.

Masotti, Louis H., "Patterns of White and Nonwhite School Referenda," *Educating an Urban Population,* Marilyn Grittle (ed.), Sage, Beverly Hills, California, 1967.

Matthews, Donald R., and James W. Prothro, *Negroes and the New Southern Politics,* Harcourt, Brace and World, New York, 1966.

McCarty, Donald J., "School Board Membership: Why Do Citizens Serve?," *Administrators Notebook,* Vol. 8, pp. 1-4, September 1959.

McDill, Edward L., and Jeanne Clare Ridley, "Status, Anomie, Political Alienation and Political Participation," *American Journal of Sociology,* Vol. 68, pp. 205-213, September 1962.

McNeal, Julia, and Margaret Rogers (eds.), *The Multi-Racial School,* Penguin Books, New York, 1972.

McPartland, James M., and Joyce A. Epstein, "Interim Report: School Organization and Student Outcomes—A Study of the Effects of Open-Environment Schools," Report 166, Center for Social Organization of Schools, The Johns Hopkins University, Baltimore, Maryland, December 1973.

Medley, Donald M., and Harold E. Mitzel, "A Technique for Measuring Classroom Behavior," *Journal of Educational Psychology,* Vol. 49, No. 2, pp. 86-92, 1958.

——, "Some Behavioral Correlates of Teacher Effectiveness," *Journal of Educational Psychology,* Vol. 50, No. 6, pp. 239-246, December 1959.

——, "A Tentative Framework for the Study of Effective Teacher Behavior," *Journal of Experimental Education, Vol. 30, No. 4, pp. 317-320, June 1962.*

Meeker, Robert J., and Daniel M. Weiler, *A New School for the Cities,* System Development Corporation, Santa Monica, California, August 1970.

Mendoza, Sonia M., Thomas L. Good, and Jere E. Brophy, "The Communication of Teacher Expectations in a Junior High School," paper presented at the 1971 meeting of the American Educational Research Association, New York, 1971.

Miles, Matthew B. (ed.), *Innovation in Education,* New York Teachers College, Columbia, 1964.

Miller, Arthur H., et al., "A Majority Party in Disarray: Policy Polarization in the 1972 Election," paper presented at the Annual Meeting of the American Political Science Association, September 1973.

Miller, Arthur H., Thad A. Brown, and Alden S. Raine, "Social Conflict and Political Estrangement, 1958-1972," paper presented at the Annual Meeting of the Midwest Political Science Association, May 1973.

Miller, James, "Early Childhood Education as an Intervention in the Child's Ecological System," National Laboratory for Early Childhood Education (Xeroxed report).

Miller, Warren E., and Donald E. Stokes, "Constituency Influence in Congress," *American Political Science Review,* Vol. 57, pp. 45-56, March 1963.

Mills, C. Wright, and Melville J. Ulmer, "Small Business and Civil Welfare," *The Structure of Community Power,* Michael Aiken and Paul E. Mott (eds.), Random House, New York, 1970.

Mills, Nicolaus (ed.), *The Great School Bus Controversy,* Teachers College Press, New York, November 1973.

Minar, David W., "The Community Basis of Conflict in School System Politics," *American Sociological Review,* Vol. 31, pp. 822-835, December 1966.

Minuchin, P., et al., *The Psychological Impact of School Experience,* Basic Books, New York, 1969.

Modiano, N., "Bilingual Education for Children of Linguistic Minorities," *American Indigena,* Vol. 28, pp. 405-414, 1968.

Morlock, Laura, *Black Power and Black Influence in 91 Northern Cities,* Ph.D. Dissertation, The Johns Hopkins University, 1973.

——, "Business Interests, Countervailing Groups and the Balance of Influence in 91 Cities," *The Search for Community Power,* 2nd ed., Willis D. Hawley and Frederick M. Wirt (eds.), Prentice-Hall, Englewood Cliffs, New Jersey, 1973.

Morrison, Arnold, and Donald McIntyre, *Schools and Socialization,* Penguin Books, New York, 1971.

Mosteller, Frederick, and Daniel P. Moynihan (eds.), *On Equality of Educational Opportunity,* Vintage Books, New York, 1972.

Moynihan, Daniel P., *The Negro Family: The Case for National Action,* U.S. Department of Labor, Washington, D.C., 1965.

Muir, William K., Jr., *Prayer in Public Schools: Law and Attitude Change,* The University of Chicago Press, Chicago, 1967.

Murphy, Walter F., and Joseph Tanenhaus, "Public Opinion and the Supreme Court: The Goldwater Campaign," *Public Opinion Quarterly,* Vol. 32, pp. 32-50, Spring 1968.

Muse, Benjamin, *The American Negro Revolution,* Citadel Press, New York, 1968.

——, *Ten Years of Prelude: The Story of Integration Since the Supreme Court's 1954 Decision,* Viking Press, New York, 1964.

——, *Virginia's Massive Resistance,* Indiana University Press, Bloomington, 1961.

National Commission for the Defense of Democracy Through Education, *Indianapolis, Indiana: A Study of the Sudden Forced Resignation of a Superintendent,* National Education Association, Washington, D.C., 1960.

National Opinion Research Center, *Southern Schools: An Evaluation of the Effects of the Emergency School Assistance Program and of School Desegregation—Vols. I and II,* NORC Report 124B, Chicago, 1973.

Nie, Norman, Dale H. Bent, and C. Hadlai Hull, *Statistical Package for the Social Sciences,* McGraw-Hill, New York, 1970.

Noar, Gertrude, *The Teacher and Integration,* Student National Education Association, Washington, D.C., 1966.

Oppenheim, Abraham N., *Questionnaire Design and Attitude Measurement,* Basic Books, New York, 1966.

Patterson, H. D., "The Errors of Lattice Sampling," *Journal of the Royal Statistical Society,* Vol. 16, No. 1, pp. 140-149, 1954.

Pettigrew, Thomas F., "Demographic Correlates of Border-State Desegregation," *American Sociological Review,* Vol. 22, pp. 683-689, December 1957.

——, *Racially Separate or Together,* McGraw-Hill, New York, 1971.

——, "Regional Differences in Anti-Negro Prejudice," *Journal of Abnormal and Social Psychology,* Vol. 59, pp. 28-36, July 1959.

——, "When a Black Candidate Runs for Mayor: Race and Voting Behavior," *People and Politics in Urban Society,* Harlan Hahn (ed.), Sage, Beverly Hills, 1972.

Pettigrew, Thomas F., and Robert Riley, "Contextual Models of School Desegregation," *Attitudes, Conflicts and Social Change,* Bert King and Elliot McGinnies (eds.), Academic Press, New York, 1972.

Pettigrew, Thomas F., and Michael J. Ross, *Understanding Racial Change,* Harvard University Press, Cambridge, forthcoming.

Pettigrew, Thomas F., Elizabeth L. Useem, Clarence Normand, and Marshall S. Smith, "Busing: A Review of 'The Evidence'," *Public Interest,* Vol. 30, pp. 88-118, Winter 1973.

Piele, Philip K., *School Financial Elections, Educational Management Review Series, No. 3,* National Center for Educational Research and Development, Washington, D.C., February 1972.

Pinard, Maurice, "Structural Attachments and Political Support in Urban Politics: The Case of Fluoridation Referenda," *American Journal of Sociology,* Vol. 68, pp. 513-526, 1963.

Plaut, Thomas A., "Analysis of Voting Behavior in a Fluoridation Referenda," *Public Opinion Quarterly,* Vol. 23, pp. 213-222, Summer 1959.

Pomper, Gerald M., "From Confusion to Clarity: Issues and American Voters, 1956-1968," *American Political Science Review,* Vol. 66, pp. 415-428, June 1972.

Porter, Judith D. R., *Black Child, White Child: The Development of Racial Attitudes,* Harvard University Press, Cambridge, Massachusetts, 1971.

Presthus, Robert, *Men at the Top: A Study in Community Power,* Oxford University Press, Oxford, 1968.

Proshansky, Harold, and Peggy Newton, "The Nature and Meaning of Negro Self-Identity," *Social Class, Race and Psychological Development,* Martin Deutsch, Irwin Katz, and Arthur R. Jensen (eds.), Holt, Rinehart and Winston, New York, 1968.

Ramos, M., J. V. Aguilar, and B. F. Sibayan, "The Determination and Implementation of Language Policy," Philippine Center for Language Study Monograph Series No. 2, Alemar/Phoenix, Quezon City, Philippines, 1967.

Rappoport, Michael, "The Distinction Pollsters Don't Make," *Washington Monthly,* pp. 13-15, March 1974.

Reagan, Michael V., et al., *Busing: Ground Zero in School Desegregation: A Literature Review with Policy Recommendations,* Syracuse University Research Corporation, Syracuse, New York, 1972.

Reed, John S., *The Enduring South: Subcultural Persistence in Mass Society,* D. C. Heath, Lexington, Massachusetts, 1972.

Rist, Ray C., *The Urban School: A Factory for Failure,* MIT Press, Cambridge, Massachusetts, 1973.

——, "Student Social Class and Teacher Expectations: The Self-Fulfilling Prophecy in Ghetto Education," *Harvard Educational Review,* Vol. 41, No. 3, August 1970.

Roberts, S. A., and Cassell P. Horton, "Extent of and Effects of School Desegregation," *Comparative Studies of Blacks and Whites in the United States,* Kent S. Miller and Ralph M. Dreger (eds.), Seminar Press, New York, 1973.

Robinson, John P., "Public Reaction to Political Protest: Chicago, 1968," *Public Opinion Quarterly,* Vol. 34, pp. 1-9, Spring 1970.

Robinson, William S., "Ecological Correlations and the Behavior of Individuals," *American Sociological Review,* Vol. 15, pp. 351-357, June 1950.

Rodgers, Harrel R., and Charles S. Bullock, III, *Law and Social Change: Civil Rights Laws and Their Consequences,* McGraw-Hill, New York, 1972.

Rogers, David, *110 Livingston Street: Politics and Bureaucracy in the New York City School System,* Random House, New York, 1968.

Rokeach, Milton, *Beliefs, Attitudes, and Values,* Jossey-Bass, San Francisco, 1968.

Rosenthal, R., and Lenore Jacobson, "Self-Fulfilling Prophecies in the Classroom: Teachers' Expectations as Maintained Determinants of Pupils' Intellectual

Competence," *Social Class, Race and Psychological Development,* Martin Deutsch, Irwin Katz, and Arthur Jensen (eds.), Holt, Rinehart and Winston, New York, 1968.

Rosenshine, Barak, and Norma Furst, "Research on Teacher Performance Criteria," *Research in Teacher Education: A Symposium,* B. O. Smith (ed.), Prentice-Hall, New Jersey, 1971.

Rosenshine, Barak, *Teaching Behaviors and Student Achievement,* National Foundation for Educational Research, England, 1971.

Ross, J. Michael, *Racial Change in the Urban North: 1962-1968,* D. C. Heath, Lexington, Massachusetts, forthcoming.

——, "Resistance to Racial Change in the Urban North: 1962-1968," Ph.D. Dissertation, Harvard University, 1973.

Ross, J. Michael, T. F. Pettigrew, and Thomas Crawford, "Negro Neighbors—Banned in Boston," *Transaction,* Vol. 3, pp. 13-18, September-October 1966.

Rossell, Christine H., "Measuring School Desegregation," Chapter 12, in *Political Strategies in Northern School Desegregation,* by David J. Kirby, T. Robert Harris, and Robert L. Crain, D. C. Heath, New York, 1973.

Rubin, Lillian B., *Busing and Backlash: White Against White in an Urban School District,* University of California Press, Berkeley, 1972.

Runciman, Walter G., *Relative Deprivation and Social Justice,* University of California Press, Berkeley, 1966.

Salinas, Guadalupe, "Mexican-Americans and the Desegregation of Schools in the Southwest," *Houston Law Review,* Vol. 8, No. 5, pp. 929-951, May 1971.

Sarratt, Reed, *The Ordeal of Desegregation: The First Decade,* Harper and Row, New York, 1966.

St. John, Nancy, "The Effect of Segregation on the Aspirations of Negro Youth," *Harvard Educational Review,* Vol. 36, No. 3, Summer 1966.

——, "Social Psychological Aspects of School Desegregation," *Rethinking Urban Education,* Herbert J. Walberg and Andrew T. Kopan (eds.), Jossey-Bass, San Francisco, 1972.

——, "Thirty-Six Teachers: Their Characteristics and Outcomes for Black and White Pupils," *American Educational Research Journal,* Vol. 8, No. 4, pp. 635-648, November 1971.

Sarason, Seymour B., *The Culture of the School and the Problem of Change,* Allyn and Bacon, Boston, 1971.

Sarason, Seymour B., et al., *Anxiety in Elementary School Children,* John Wiley and Sons, New York, 1960.

Schneider, E. Joseph, and Mary Kennedy Burton, "Inner City Teaching: A Success Story," *The Urban Review,* Vol. 6, No. 4, pp. 28-31, 1973.

Schrag, Peter, *Village School Downtown: Boston Schools, Boston Politics,* Beacon Press, Boston, 1967.

Schuman, Howard, and Barry Gruenberg, "Dissatisfaction with City Services: Is Race an Important Factor?," *People and Politics in Urban Society,* Harlan Hahn (ed.), Sage, Beverly Hills, 1972.

Schuman, Howard, "The Impact of City on Racial Attitudes," *American Journal of Sociology,* Vol. 76, pp. 213-261, September 1970.

Schwartz, Audrey James, "The Social Context of the Integrated School," paper presented at the Pacific Sociological Association Meeting, May 1973.

Schwartz, Mildred A., *Trends in White Attitudes Toward Negroes,* National Opinion Research Center, Chicago, 1967.

Seashore, Stanley E., and D. G. Bowers, *Changing the Structure and Functioning of an Organization,* University of Michigan Press, Survey Research Center, Ann Arbor, Monograph No. 33, 1963.

Seeman, Melvin, "On the Meaning of Alienation," *American Sociological Review,* Vol. 24, pp. 783-790, December 1959.

Sheatsley, Paul, "White Attitudes Toward the Negro," *Daedalus,* Vol. 95, pp. 217-238, Winter 1966.

Shapiro, Michael J., "Rational Political Man: A Synthesis of Economic and Socio-Psychological Perspectives," *American Political Science Review,* Vol. 65, pp. 1106-1119, December 1969.

Shively, W. Philips, "Ecological Inference: The Use of Aggregate Data to Study Individuals," *American Political Science Review,* December 1969.

Shoemaker, Don (ed.), *With All Deliberate Speed: Segregation-Desegregation in Southern Schools,* Harper and Brothers, New York, 1957.

Skipper, James L., Jr., Anthony L. Guenther, and Gilbert Nass, "The Sacredness of .05: A Note Concerning the Uses of Statistical Levels of Significance in Social Science," *The American Sociologist,* Vol. 2, pp. 16-19, February 1967.

Smith, Al, Anthony Downs, and M. Leanne Lachman, *Achieving Effective Desegregation,* Real Estate Research Corporation, D. C. Heath, Lexington, Massachusetts, 1973.

Smith, Bob, *They Closed Their Schools: Prince Edward County, Virginia, 1951-1964,* University of North Carolina Press, Chapel Hill, 1965.

Smith, Lewis M., and Pat M. Kieth, *Anatomy of Educational Innovation: An Organizational Analysis of an Elementary School,* John Wiley and Sons, New York, 1971.

Smith, Marshall S., "Equality of Educational Opportunity: The Basic Findings Reconsidered," *On Equality of Educational Opportunity,* Frederick Mosteller and Daniel P. Moynihan (eds.), Vintage Books, New York, 1972.

Snow, Richard E., and Janet D. Elashoff, *Pygmalion Reconsidered,* Charles A. Jones, Worthington, Ohio, 1971.

Stein, A., "Strategies for Failure," *Harvard Educational Review,* Vol. 61, No. 2, May 1971.

Stone, Clarence, "Local Referendums: An Alternative to the Alienated Voter Model," *Public Opinion Quarterly,* Vol. 29, pp. 213-222, Summer 1965.

Stone, I. F., "Moving the Constitution to the Back of the Bus," *New York Review of Books,* Vol. 18, pp. 4-11, April 20, 1972.

Stryker, John, "Representation and Policy-Making: School Desegregation in Berkeley and Pasadena," paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, April 27-29, 1972.

Sullivan, Neil V., "Desegregation Techniques," *Education Parks,* Clearinghouse Publications, No. 9, U.S. Government Printing Office, Washington, D.C., October 1967.

——, "Integrating the Desegregated School," *Developments in School Desegregation and the Law,* University of Michigan School of Education, Ann Arbor, 1972.

Taeuber, Karl A., and Alma F. Taeuber, *Negroes in Cities,* Aldine, Chicago, 1965.

Teele, James E., *Evaluating School Busing: Case Study of Boston's Operation Exodus,* Praeger, New York, 1973.

——, *Evaluating School Busing,* Praeger, New York, 1973.

Tepping, B. J., W. H. Hurwitz, and W. E. Deming, "On the Efficiency of Deep Stratification in Block Sampling," *Journal of the American Statistical Association,* Vol. 38, pp. 93-100, 1943.

Tiedt, Sidney W., "Oregon School Board Members in the Willamette Valley," *Oregon School Study Council Bulletin,* Vol. 6, University of Oregon, School of Education, Eugene, 1962.

Tumin, Melvin M., *Desegregation: Resistance and Readiness,* Princeton University Press, Princeton, New Jersey, 1958.

U.S. Commission on Civil Rights, *Education Parks: Appraisals of Plans to Improve Educational Quality and Desegregate the Schools,* Clearinghouse Publications, No. 9, U.S. Government Printing Office, Washington, D.C. October 1967.

——, *Racial Isolation in the Public Schools,* U.S. Government Printing Office, Washington, D.C., 1967.

——, *Teachers and Students: Report V: Mexican-American Education Study,* U.S. Government Printing Office, Washington, D.C., 1973.

——, *School Desegregation in Ten Communities,* U.S. Government Printing Office, Washington, D.C., June 1973.

——, *Racial Isolation in the Public Schools,* U.S. Government Printing Office, Washington, D.C., 1967.

U.S. Congress, House, Committee on the Judiciary Subcommittee No. 5, *School Busing,* Part I, 92nd Cong., 2nd sess., February 28, 29; March 3, 4, 10, 18, and 24, 1972, U.S. Government Printing Office, Washington, D.C., 1972.

U.S. Department of Health, Education and Welfare, *Directory of Public Elementary and Secondary Schools in Selected Districts, Fall 1970 Enrollment and Staff by Racial/Ethnic Group,* U.S. Government Printing Office, Washington, D.C., 1971.

U.S. National Center for Educational Statistics, Office of Education, *Bond Sales for Public Schools Purposes,* U.S. Government Printing Office, Washington, D.C. (annual reports).

U.S. Office of Education, *Planning Educational Change—Vol. IV: How Five School Systems Desegregated,* U.S. Government Printing Office, Washington, D.C., 1969.

——, "Evaluation of the Emergency School Aid Act (ESAA) Basic LEA Program: A Comparative Evaluation of Three Types of Educational Intervention," RFP 73-29, 1973.

Valentine, C., "Deficit, Difference and Bicultural Models of Afro-American Behavior," *Harvard Educational Review,* Vol. 41, No. 2, May 1971.

——, *Culture and Poverty: Critique and Counter Proposals,* The University of Chicago Press, Chicago, 1968.

Vanneman, Reeve D., and Thomas F. Pettigrew, "Race and Relative Deprivation in the Urban United States," *Race,* Vol. 13, pp. 461-486, April 1972.

Verba, Sidney, and Norman Nie, *Participation in American Politics,* Random House, New York, 1973.

Walberg, Herbert J., "Social Environment as a Mediator of Classroom Learning," *Journal of Educational Psychology,* Vol. 60, pp. 443-448, 1969.

Wall, Marvin, "What the Public Doesn't Know Hurts," *Civil Rights Digest,* Summer 1973.

Wasby, Stephen L., *The Impact of the United States Supreme Court: Some Perspectives,* Dorsey Press, Homewood, Illinois, 1970.

Weinberg, Meyer, *Desegregation Research: An Appraisal,* Phi Delta Kappa, Bloomington, Indiana, 1970.

Westie, Frank R., "Negro-White Status Differentials and Social Distance," *American Sociological Review,* Vol. 17, pp. 550-558, October 1952.

——, "A Technique for the Measurement of Race Attitudes," *American Sociological Review,* Vol. 18, pp. 73-78, February 1953.

Westie, Frank R., and David H. Howard, "Social Status Differentials and the Race Attitudes of Negroes," *American Sociological Review,* Vol. 19, pp. 584-591, October 1954.

Wilensky, Harold L., and Hugh Edwards, "The Skidder: Ideological Adjustments of Downward Mobile Workers," *American Sociological Review,* Vol. 24, pp. 215-231, April 1959.

Williams, Ben, "Achieving Effective Desegregation: A Handbook by Anthony Downs and M. Leanne Leachman," *Developments in School Desegregation and the Law,* Charles Moody, Charles Vergon, and John Taylor (eds.), University of Michigan School of Education, Ann Arbor, 1972.

Williams, Robin M., Jr., and Margaret W. Ryan (eds.), *Schools in Transition: Community Experiences in Desegregation,* University of North Carolina Press, Chapel Hill, 1954.

Willis, Charles L., "An Analysis of Voter Response to School Financial Proposals," *Public Opinion Quarterly,* Vol. 31, pp. 648-651, Winter 1967-68.

Winer, Ben J., *Statistical Principles in Experimental Design,* McGraw-Hill, New York, 1962.

Wirt, Frederick M., and Michael W. Kirst, *The Political Web of American Schools,* Little, Brown and Company, Boston, 1972.

Wolfinger, Raymond, and John Osgood Field, "Political Ethos and the Structure of City Government," *American Political Science Review,* Vol. 60, pp. 306-326, June 1966.

Wood, Robert, "Response-Contingent Testing," *Review of Educational Research,* Vol. 43, pp. 529-544, February 1973.

Yates, Frank, *Sampling for Censuses and Surveys,* 2nd rev. ed., Charles Griffin and Company, London, 1953.

Ziegler, Harmon, *The Political Life of American Teachers,* Prentice-Hall, Englewood Cliffs, New Jersey, 1967.

Zukosky, Jerome, "Giving Up On Integration," *The New Republic,* pp. 19-21, October 14, 1972.