DOCUMENT RESUME

ED 271 810 EA 018 552

AUTHOR McLaughlin, Donald H.; Wise, Lauress L.
TITLE Nonpublic Education of the Nation's Children.
INSTITUTION American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY National Center for Education Statistics (DHEW), Washington, D.C.
REPORT NO AIR-SAGE-TR-9
PUB DATE Apr 80
CONTRACT 300-78-0150
NOTE 107p.; Prepared by the Statistical Analysis Group in Education.
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160) -- Statistical Data (110)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Longitudinal Studies; National Surveys; Parochial Schools; *Private Education; Questionnaires; Religious Factors; School Demography; *School Statistics; Tables (Data); Validity
IDENTIFIERS National Center for Education Statistics; Nonpublic School Survey (NCES)

ABSTRACT

Data gathered in a 3-year study of nonpublic education are presented and discussed in this report. The nation's 20,000 nonpublic elementary and secondary schools were all asked to complete the Nonpublic School Survey in the school years 1976-77, 1977-78, and 1978-79. Items in the survey covered directory information (names and addresses), size of town, mix of boarding and day students, mix of male and female students, school type (elementary, vocational, etc.), types of curriculum offered, nonlanguage courses taught in foreign languages, receipt of federal program funds, religious affiliation, numbers of graduates (by sex), enrollment (by grade), teachers (religious or lay) and other staff by sex, and income and expenditure information. Data from the survey can be used to address economic issues (related to increases or decreases in the nonpublic share of the educational market) and service issues (related to the quality and breadth of services provided by nonpublic educational institutions). The data permit assessment of short-term trends and identification of religious, regional, and other differences among schools. Appendices, which make up about two-thirds of the report, discuss plans for further surveys, present the survey form, describe the merging of files from all three years, and provide a detailed account of procedures used for imputing missing data. (PGD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED271810

Technical Report No. 9

Nonpublic Education of the Nation's Children

Donald H. McLaughlin
Laurese L. Wise

Prepared by

SAGE
STATISTICAL ANALYSIS GROUP IN EDUCATION

For the

National Center for Education Statistics

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

EA 018 552

American Institutes for Research
Box 1113, Palo Alto, California 94302

TECHNICAL REPORT NO. 9

NONPUBLIC EDUCATION OF THE NATION'S CHILDREN

Donald H. McLaughlin

Lauress L. Wise

Statistical Analysis Group in Education
American Institutes for Research
P.O. Box 1113
Palo Alto, California 94302

This work was done under Contract No. 300-78-0150 with the National Center for Education Statistics, Department of Health, Education, and Welfare. However, the content does not necessarily reflect the position or policy of either agency, and no official endorsement should be inferred.

April 1980

FOREWORD

This report was prepared by the Statistical Analysis Group in Education (SAGE) for the National Center for Education Statistics (NCES). It focuses on the interpretation of the results of the Nonpublic School Survey as they relate to current issues in federal policy for nonpublic schools. This survey, although limited in scope of content, contributes information that sheds light on a number of policy alternatives. The major strengths of this survey are (1) that it establishes a data base on the universe, not a sample, of nonpublic schools and (2) that data are available for three successive years, thus enabling development of short-term trend projections.

The surveys were carried out by the National Catholic Education Association (NCEA) and the Council for American Private Education (CAPE) under authorization of the Institutional Surveys Branch of the Division of Elementary and Secondary Education at NCES. Rhoda Goldstein directed the effort at NCEA, and Robert Lamborn directed the effort at CAPE. Joanell Porter served as Project Officer for NCES. The development of the present report was carried out under the guidance of Roy Nehrt, Richard Cook, and Rolf Wulfsberg, and of the Project Officer for SAGE, Roslyn Korb. The statistical imputation of missing data, the major contribution of SAGE to these analyses, was reviewed at an early stage by Donald Rubin and William Madow. Files for use by SAGE were prepared at NCES by Ted Chmura and Frank Morgan; and Paul Dunn and other staff at NCES and at the American Institutes for Research facilitated production of this report. The authors are grateful to them all.

The main body of this report, which is relatively short, presents substantive results of the surveys as they relate to issues. The report is supplemented by four appendixes, the first of which considers the implications of issues not addressed by this survey for future data needs and recommends a comprehensive data base development plan for nonpublic school information. The survey questionnaire is presented as a second

appendi , and the results of merging the three years' files are presented in a third appendix. The fourth appendix documents the statistical imputation of missing data carried out by SAGE and reports the results of a validation study of the imputation processes.

D. H. McLaughlin
L. L. Wise
30 June 1980

TABLE OF CONTENTS

	Page No.
Foreword	
Introduction	1
Results	4
General Results	4
Economically Related Results	9
Service Related Results	12
Community Size and Region	12
Sex of Students Served	18
Day/Boarding Facilities	23
Federal Program Participation	28
References	31
APPENDIX A - Future Survey Plans	33
APPENDIX B - Nonpublic School Survey Form	51
APPENDIX C - File Development and Merging	55
APPENDIX D - Description and Validation of the Missing Data Imputation Procedures	61

List of Exhibits

Exhibit		Page No.
A	Estimated Number of Nonpublic School Openings and Closings, by Religious Affiliation	9
B	Percentage of Closures by Size of School Prior to Closing	10
C	Percentages of Schools Newly Opening and Newly Closing, by Region	13

List of Tables

Table		Page No.
1	Number of Schools by Religious Affiliation and School Type	6
2	Number of Teachers by Religious Affiliation and School Type	6
3	Number of Students (in Thousands) by Religious Affiliation and School Type	7
4	Current Operating Expenditures (in Millions) by Religious Affiliation and School Type	7
5	Pupil/Teacher Ratio by Religious Affiliation and School Type	8
6	Per-Pupil Expenditures by Religious Affiliation and School Type	11
7	Number of Graduates by Religious Affiliation and School Type	11
8	Number of Schools by Religious Affiliation and Community Size	14
9	Number of Schools by School Type and Community Size	14
10	Number of Students (in thousands) by Religious Affiliation and Community Size	15
11	Number of Students (in thousands) by School Type and Community Size	15
12	Pupil/Teacher Ratio by Religious Affiliation and Community Size	16
13	Pupil/Teacher Ratio by School Type and Community Size	16
14	Per-Pupil Expenditures by Religious Affiliation and Community Size	17
15	Per-Pupil Expenditures by School Type and Community Size	17
16	Number of Schools by Religious Affiliation and Sex of Students Served	19
17	Number of Schools by School Type and Sex of Students Served	19

List of Tables (continued)

Table		Page No.
18	Number of Students (in thousands) by Religious Affiliation and Sex of Students Served	20
19	Number of Students (in thousands) by School Type and Sex of Students Served	20
20	Pupil/Teacher Ratio by Religious Affiliation and Sex of Students Served	21
21	Pupil/Teacher Ratio by School Type and Sex of Students Served	21
22	Per-Pupil Expenditures by Religious Affiliation and Sex of Students Served	22
23	Per-Pupil Expenditures by School Type and Sex of Students Served	23
24	Number of Schools by Religious Affiliation and Day-Boarding Mix	24
25	Number of Schools by School Type and Day-Boarding Mix	24
26	Number of Students (in thousands) by Religious Affiliation and Day-Boarding Mix	25
27	Number of Students (in thousands) by School Type and Day-Boarding Mix	25
28	Pupil/Teacher Ratio by Religious Affiliation and Day-Boarding Mix	26
29	Pupil/Teacher Ratio by School Type and Day-Boarding Mix	26
30	Per-Pupil Expenditures by Religious Affiliation and Day-Boarding Mix	27
31	Per-Pupil Expenditures by School Type and Day-Boarding Mix	27
32	Number of Schools Receiving Federal Assistance, by Religious Affiliation	29
33	Number of Schools Receiving Federal Assistance, by School Type	29

NONPUBLIC EDUCATION OF THE NATION'S CHILDREN

Introduction

How many of our nation's children are being educated outside the public school system? Are private and parochial schools gaining students at the expense of public schools, or on the other hand, are they on the verge of collapse from economic pressures and declining enrollments? Do private and parochial schools represent an essential dimension of freedom of choice in American society, or on the other hand, do they represent a barrier to programs to provide equal opportunity to all young people through education? What is the public policy role in private education? In the Education Amendments of 1978 (P.L. 95-561), Congress clearly and repeatedly demonstrated its interest in the role of the nonpublic sector in elementary and secondary education, including provisions for serving the special educational needs of all children, without regard for whether they are enrolled in public or in private schools. Numerous court cases continue to arise concerning public aid to children enrolled in nonpublic schools.

Responding to the growing needs for information on the country's nonpublic schools, the National Center for Education Statistics has for three years conducted a survey of the nation's 20,000 nonpublic elementary and secondary schools. This was a largely exploratory project, involving the Council for American Private Education and the National Catholic Education Association in a cooperative survey effort directly serving the data needs of the nonpublic schools while continuing the development of a national nonpublic school data base. Alternative designs for the data base, with recommendations based on the results of these three years, can be found in Appendix A. Other significant data sources include the National Catholic Education's Data Bank and the Census Bureau's Population Surveys.

The NCES survey included the same items for each of the three school years, 1976-77, 1977-78, and 1978-79, making longitudinal analyses feasible.

On the other hand, although this was a universe survey, the breadth of information gathered was limited, thus avoiding unnecessary burden on the nonpublic school respondents. Items in the survey focus on:

1. Directory information (name, address, telephone)
2. Size of town or city
3. Boarding-day student mixture
4. Male-female student mixture
5. School type (elementary, middle, secondary, combined, vocational, special, or alternative)
6. Types of curriculum offered
7. Nonlanguage courses taught in a foreign language
8. Receipt of federal program funds (Title I, Title IV-B, Title IV-C, school lunch)
9. Religious affiliation
10. Number of graduates, by sex
11. Enrollment, by grade
12. Teachers, religious or lay, and other staff, by sex
13. Income and expenditures

The survey form is reproduced as Appendix B. Readers will note that assessment of student characteristics and of quality of instruction were not feasible in this survey. Therefore, issues requiring such data will not be addressed in this report.

The major issues that can be addressed using the data from this survey fall into two categories. Other categories that might be addressed in future surveys are discussed in Appendix A.

One category of issues is economic, related to the impact of increases or decreases in the nonpublic share of the educational market. To address these issues, for which Congress has recently mandated a study (P.L. 94-561, Section 1203(e)(10)), basic statistics enabling projections of enrollment, staffing, and expenditures in the nonpublic sector will be necessary. Specifically, in that Section, Congress called for "an analysis of current and future Federal assistance for nonpublic elementary and secondary edu-

cation, including the extent of nonpublic participation in Federal programs, trends in enrollments and costs of private education, the impact of private schools on public school enrollments and financial support, and an examination of alternative Federal policies for support of private education." The universe data provided in the NCES survey is designed to complement indepth studies of samples of schools in contributing to policies concerning public support of nonpublic schools.

The other category of issues to be discussed concern service: the types and quality of nonpublic education opportunities available. These issues deal with the controversy of whether and to what extent nonpublic schools offer programs that are either inferior or superior to programs offered in the public schools, both in breadth of opportunity and in efficiency and effectiveness. Such issues are likely to increase as use of competency examination spreads as a criterion for student and school performance. For example, the State Supreme Court in Kentucky recently ruled that private schools in that state are not bound to meet the same course, teacher, and textbook accreditation standards as public schools but may be monitored for achievement through a standardized testing program. To resolve questions of efficiency and effectiveness requires much more indepth information than the NCES surveys provide; however, two proxy measures are provided, measures used in other studies and for which relations to efficiency and effectiveness must be ultimately clarified: per-pupil expenditures and pupil/teacher ratios. Expenditures, it must be noted, are the least trustworthy data in the NCES surveys, because many nonpublic school administrators were not easily able to fit their financial statements into the survey categories. This led to a large portion of nonresponse (about 50%), and it also brings into question the meaningfulness of the responses that were obtained. Nevertheless, values were imputed for missing data, and data on per-pupil current operating expenditures are included in this report.

Concerning breadth of services, there are data on the number of nonpublic schools of various types and affiliations that provide boarding facilities, that limit contact with the opposite sex in the classroom, that

provide vocational and special education, that teach some (nonlanguage) courses in a foreign language, that provide services for educationally disadvantaged children, and that provide services in areas of varying population density. These and other current issues were discussed at a Seminar on Meaning and Significance in the Analysis of the Nonpublic School Survey (McLaughlin, 1978).

Specifically, the data from these surveys provide the basis for (1) estimating the level and short-term projections of children attending and graduating from, teachers teaching in, and funds being expended in nonpublic schools; (2) identifying factors predictive of nonpublic school closures; (3) estimating the level and trends in per-pupil expenditures in different types of schools; (4) comparing pupil/teacher ratios among schools; (5) describing differences, such as in enrollment trends, among schools with varying religious affiliations; (6) estimating nonpublic school participation in federal programs; and (7) describing geographic distributions of nonpublic schools. The data unfortunately are not adequate to estimate the number of openings of nonpublic schools each year with confidence, because the coverage of newly opened schools by the survey has not been adequately verified.

The methods used in developing the three-year longitudinal file of data from the nonpublic school survey program are described in detail in Appendix C. A computer file with documentation, including indicators of which data were imputed by SAGE and which were actually produced by nonpublic school respondents, is available for further analyses. The Data Service Branch of NCES can be contacted for copies of the file and documentation.

Results

General Results

The figures in Tables 1, 2, 3, and 4 provide an overview of the nonpublic elementary and secondary schools over the period from 1976 to 1978. From Table 1, it can be seen that there were about 20,000 nonpublic schools

in each of the last three school years.* Compared with the public school system's 87,365 schools in 1978, this represents a substantial portion of the nation's educational establishment. Nonpublic schools are particularly prevalent at the elementary and middle levels, where they account for more than 20% of the schools. According to Table 2, however, counts of classroom teachers indicate that only 11% of the nation's 2,500,000 teachers are in nonpublic schools. Clearly, individual nonpublic schools tend to have fewer teachers than public schools, and as seen by comparing Table 3 with public school enrollment figures, fewer students. For 1976-77, 1977-78, and 1978-79, 11.7%, 11.8%, and 12.0%, respectively, of the nation's elementary and secondary students were in nonpublic schools. While enrollment in Lutheran schools and in non-affiliated schools remained nearly constant, enrollment in Catholic schools declined 4% over the two-year period, and enrollment in schools with other religious affiliations rose by 10%. Clearly statements about increasing or decreasing enrollments should not be made for nonpublic schools in general but must be made separately for different categories of schools.

Current operating expenditures were reported in each survey for the preceding school year, and the results for the second and third surveys are shown in Table 4. All expenditure and income data from this survey should be considered as very tentative, however, due to the large percentage of nonresponders, whose data were imputed statistically. Nevertheless, we estimate that about $5 billion is spent annually by nonpublic schools for current operations and an additional $0.5 billion for auxiliary operations,

---

*This estimate is based on a particular definition of school, excluding nursery schools not extending through grade 2, day care centers, and various special purpose institutions. It also depends critically on the success of the methodology for locating new schools when they open.

Results published by the National Catholic Education Association (Bredeweg, 1979a, 1979b) corroborate these figures for Catholic schools except for small differences relating to definitions of school type. NCEA published data on elementary and secondary schools indicating in 1977-78 a total of 9758 schools, with 150,333 teachers, serving 3,289,000 students, with a total operating expenditure of $1,986 million.

TABLE 1. NUMBER OF SCHOOLS BY RELIGIOUS AFFILIATION AND SCHOOL TYPE

SCHOOL TYPE	RELIGIOUS AFFILIATION									
	Catholic		Lutheran		Other Affiliations		No Affiliation		TOTAL	
	Count	Pct Chng	Count	Pct Chng	Count	Pct Chng	Count	Pct Chng	Count	Pct Chng
Elementary and Middle										
1976/1977	8092		1358		2039		1476		12965	
1977/1978	8054	-0	1356	-0	2071	2	1453	-2	12934	-0
1978/1979	8020	-0	1345	-1	200	-3	1384	-5	12749	-1
Secondary										
1976/1977	1552		71		388		473		2484	
1977/1978	1538	-1	74	4	382	-2	463	-1	2462	-1
1978/1979	1518	-1	74	0	381	-0	445	-5	2418	-2
Combined Elementary and Secondary										
1976/1977	140		52		1954		1273		3422	
1977/1978	138	-1	53	2	2014	3	1262	-1	3467	1
1978/1979	135	-2	51	-4	1946	-3	1219	-3	3351	-3
Other: Special, Vocational or Alternative										
1976/1977	182		14		70		946		1212	
1977/1978	183	1	15	7	68	-3	944	-0	1210	-0
1978/1979	176	-4	15	0	61	-10	896	-5	1148	-5
TOTAL										
1976/1977	9966		1495		4451		4171		20083	
1977/1978	9913	-1	1498	0	4535	2	4127	-1	20073	-0
1978/1979	9849	-1	1485	-1	4388	-3	3944	-4	19666	-2

TABLE 2. NUMBER OF TEACHERS BY RELIGIOUS AFFILIATION AND SCHOOL TYPE

Elementa and Middle										
1976/1977	95031		7872		16338		14066		133307	
1977/1978	94734	-0	8184	4	17093	5	14572	4	134583	1
1978/1979	93028	-2	8226	1	17989	5	13783	-5	133031	-1
Secondary										
1976/1977	45754		1189		6092		9086		62121	
1977/1978	45458	-1	1373	15	6404	5	9650	6	62885	1
1978/1979	44940	-1	1341	-2	7092	11	9231	-4	62604	-0
Combined Elementary and Secondary										
1976/1977	3189		531		27589		29946		61255	
1977/1978	3315	4	582	10	30877	12	31898	7	66672	9
1978/1979	2960	-11	533	-8	30380	-2	30652	-4	64525	-3
Other: Special, Vocational or Alternative										
1976/1977	2277		108		618		9233		12236	
1977/1978	2387	5	103	0	640	4	10886	18	14021	15
1978/1979	2424	2	99	-8	497	-22	9561	-12	12581	-10
TOTAL										
1976/1977	146251		9700		50637		62331		268919	
1977/1978	145894	-0	10247	6	55014	9	67006	8	278161	3
1978/1979	143352	-2	10199	-0	55958	2	63232	-6	272741	-2

TABLE 3. NUMBER OF STUDENTS(in Thousands) BY RELIGIOUS AFFILIATION AND SCHOOL TYPE

SCHOOL TYPE	RELIGIOUS AFFILIATION									
	Catholic		Lutheran		Other Affiliations		No Affiliation		TOTAL	
	Count	Pct Chng	Count	Pct Chng	Count	Pct Chng	Count	Pct Chng	Count	Pct Chng
Elementary and Middle										
1976/1977	2464		179		253		184		3081	
1977/1978	2400	-3	180	0	256	1	190	3	3025	-2
1978/1979	2342	-2	181	1	281	10	184	-3	2989	-1
Secondary										
1976/1977	882		21		79		97		1080	
1977/1978	865	-2	24	14	77	-3	98	1	1064	-1
1978/1979	857	-1	24	-3	88	15	100	2	1069	0
Combined Elementary and Secondary										
1976/1977	58		11		435		401		905	
1977/1978	58	-0	12	6	476	9	407	1	952	5
1978/1979	53	-9	12	-1	478	0	396	-3	938	-1
Other: Special, Vocational or Alternative										
1976/1977	19		1		6		76		101	
1977/1978	18	-2	1	25	5	-11	73	-3	93	-3
1978/1979	18	-1	1	26	5	-8	66	-10	90	-8
TOTAL										
1976/1977	3423		212		773		758		5167	
1977/1978	3341	-2	217	2	814	5	768	1	5140	-1
1978/1979	3270	-2	217	0	852	5	747	-3	5086	-1

TABLE 4. CURRENT OPERATING EXPENDITURES(in Millions) BY RELIGIOUS AFFILIATION AND SCHOOL TYPE

Elementary and Middle										
1976/1977	1089		114		330		358		1892	
1977/1978	1145	5	117	2	338	2	354	-1	1954	3
Secondary										
1976/1977	777		34		199		339		1349	
1977/1978	805	4	33	-4	200	0	351	4	1389	3
Combined Elementary and Secondary										
1976/1977	50		8		564		773		1395	
1977/1978	48	-4	9	5	560	-1	788	2	1405	1
Other: Special, Vocational or Alternative										
1976/1977	61		4		22		380		467	
1977/1978	59	-2	3	-17	23	0	381	0	465	-0
TOTAL										
1976/1977	1977		160		1115		1850		5103	
1977/1978	2058	4	161	1	1120	0	1874	1	5214	2

such as dormitories. According to other responses on the survey not tabulated here, 70% of current operating expenditures are financed by tuition and fees ($3.6 billion in 1976 and $3.7 billion in 1977). The estimated 2% increase in expenditures must be interpreted in the context of inflation, which proceeded at an annual rate of 7% between these two school years. Thus, in constant dollars there was actually a decrease.

Differences among schools are apparent in Table 5, which shows average pupil/teacher ratios ranging from about 25 for Catholic elementary schools to about 8 for special education, vocational, technical, and alternative schools. Except for Catholic and Lutheran schools, the average pupil/teacher ratios tended to be lower than the aggregate public school ratios of 20.3, 19.8, and 19.4 for the three years, 1976-77, 1977-78, and 1978-79.

TABLE 5. PUPIL/TEACHER RATIO BY RELIGIOUS AFFILIATION AND SCHOOL TYPE

	RELIGIOUS AFFILIATION									
	Catholic		Lutheran		Other Affiliations		No Affiliation		TOTAL	
	Average	Pct Chng	Average	Pct Chng	Average	Pct Chng	Average	Pct Chng	Average	Pct Chng
SCHOOL TYPE										
Elementary and Middle										
1976/1977	25.9		22.8		15.5		13.1		23.1	
1977/1978	25.3	-2	22.0	-4	15.0	-3	13.0	-1	22.5	-3
1978/1979	25.2	-1	22.0	0	15.6	4	13.4	3	22.5	-0
Secondary										
1976/1977	19.3		18.0		13.0		10.7		17.4	
1977/1978	19.0	-1	17.8	-1	12.0	-8	10.2	-5	16.9	-3
1978/1979	19.1	0	17.8	-0	12.4	4	10.8	7	17.1	1
Combined Elementary and Secondary										
1976/1977	18.2		20.8		15.8		13.4		14.8	
1977/1978	17.5	-4	20.1	-3	15.4	-2	12.8	-5	14.3	-3
1978/1979	17.8	2	21.7	8	15.7	2	12.9	1	14.5	2
Other: Special, Vocational or Alternative										
1976/1977	8.2		6.7		9.2		8.2		8.2	
1977/1978	7.7	-6	8.4	25	7.9	-14	6.7	-18	7.0	-15
1978/1979	7.5	-3	11.5	37	9.4	18	6.9	3	7.2	3
TOTAL										
1976/1977	23.4		21.9		15.3		12.2		19.2	
1977/1978	22.9	-2	21.2	-3	14.8	-3	11.5	-6	18.5	-4
1978/1979	22.8	-0	21.3	1	15.2	3	11.8	3	18.6	1

Economically Related Results

These results fall into three categories: (1) competition for students, (2) per-pupil expenditures, and (3) trends in closures. Enrollment trends, which are relevant to the first concern, were discussed in the general section above. Per-pupil expenditure data appear in Table 6. As would be expected, these are inversely related to pupil/teacher ratios. They can be compared to the aggregate K-12 public school estimate of $1,627 per student, but only roughly because of the dramatic differences between elementary and secondary expenditures per pupil. Comparison of expenditures among groups is made more difficult because of widely varing financial reporting methodologies. Although no consistent increase was apparent for non-Catholic schools, the increase for Catholic schools matched the national inflation rate.

Estimated numbers of nonpublic school openings and closings are shown in Exhibit A. The estimates for openings are likely to be underestimates; since a school can be set up without elaborate clearance procedures, new schools may not be included in the NCES surveys. It is likely that the number of schools may be as much as 10% higher than the current file count. However, because most of the additional schools are small, there may be

EXHIBIT A

Estimated Number of Nonpublic School Openings and Closings, by Religious Affiliation

	Fall 1977		Fall 1978	
	Openings	Closings	Openings	Closings
Catholic	8	61	0	74
Lutheran	10	13	0	18
Other Affiliations	153	80	117	268
Non-Affiliated	131	158	29	193
TOTAL	302	312	146	553

only 1% or 2% additional students. Closings, on the other hand, are more likely to have been recorded, assuming that the schools were on the universe list prior to closing.

With these caveats in mind, one can interpret the estimates in Exhibit A as indicating that 1977 was a better year for nonpublic schools than 1978, in terms of schools' continuing in operation. The groups of schools declining most in numbers in 1978 were the non-affiliated schools and, as shown in Exhibit B, schools with fewer than 100 students.* These are the schools whose existence is most sensitive to the economic climate, in terms of parents' willingness and ability to make the monetary investment in their children's education.

As shown in Table 7, the nation's nonpublic schools produce approximately 300,000 12th grade graduates every year, a significant fraction (10%) of the total graduates entering the labor force or higher education. The quality of education provided to these young people prior to graduation is therefore of great concern to future manpower planners.

EXHIBIT B

Percentage of Closures by Size of School Prior to Closing

Enrollment	Closed after 1976-77	Closed after 1977-78
< 50	4.1%	8.8%
50 to 100	3.2%	4.1%
101 to 150	1.8%	2.4%
151 to 200	1.1%	2.2%
> 200	0.5%	1.1%

*Estimates of closings were obtained both from explicit indications on returned forms and from special telephone calls to a sample of nonrespondents by SAGE staff.

TABLE 6. PER-PUPIL EXPENDITURES BY RELIGIOUS AFFILIATION AND SCHOOL TYPE

SCHOOL TYPE	RELIGIOUS AFFILIATION									
	Catholic		Lutheran		Other Affiliations		No Affiliation		TOTAL	
	Average	Pct Chng	Average	Pct Chng	Average	Pct Chng	Average	Pct Chng	Average	Pct Chng
Elementary and Middle										
1976/1977	442.1		637.2		1303.6		1942.3		614.1	
1977/1978	477.3	8	649.9	2	1318.6	1	1863.1	-4	645.7	5
Secondary										
1976/1977	880.9		1592.5		2502.5		3488.3		1248.5	
1977/1978	931.2	6	1334.7	-16	2596.3	4	3585.5	3	1305.2	5
Combined Elementary and Secondary										
1976/1977	858.2		747.8		1296.5		1927.7		1541.4	
1977/1978	830.2	-3	740.8	-1	1178.5	-9	1936.7	0	1476.0	-4
Other: Special, Vocational or Alternative										
1976/1977	3255.0		5125.3		3936.0		5023.2		4634.2	
1977/1978	3228.6	-1	3381.4	-34	4437.6	13	5186.0	3	4762.4	3
TOTAL										
1976/1977	577.7		754.7		1442.2		2439.5		987.6	
1977/1978	616.1	7	743.5	-1	1377.0	-5	2439.4	-0	1014.4	3

TABLE 7. NUMBER OF GRADUATES BY RELIGIOUS AFFILIATION AND SCHOOL TYPE

Elementary and Middle										
1976										
1977										
1978										
Secondary										
1976	193634		4779		16607		21602		241622	
1977	199511	0	5180	8	15874	-4	20665	-4	241230	-0
1978	194562	-2	5099	-2	15897	0	20600	-0	236158	-2
Combined Elementary and Secondary										
1976	4574		126		22844		29490		57034	
1977	5417	18	144	14	22681	-1	28721	-3	56963	-0
1978	4430	-18	125	-13	22376	-1	29516	3	56447	-1
Other: Special, Vocational or Alternative										
1976	1104		0		296		1436		2836	
1977	952	-14	14		281	-5	1629	13	2876	1
1978	1268	33	3817	1	283	1	1681	3	3270	14
TOTAL										
1976	204312		4905		39747		52528		301492	
1977	205880	1	5338	9	38836	-2	51015	-3	301069	-0
1978	200260	-3	5262	-1	38556	-1	51797	2	295875	-2

Service Related Results

The next 24 tables are presented as three groups: Tables 8 through 15, dealing with the size of community served; Tables 16 through 23, dealing with the sex of students served; and Tables 24 through 31, dealing with the mix of day and boarding students. Finally, Tables 32 and 33 present frequencies of participation in federal programs by religious affiliations and by school type.

Community size and region. Distribution of nonpublic schools across community sizes is important because of differential educational needs of large cities, small cities, and rural areas. As Table 8 shows, schools of each of the major religious affiliation categories are spread similarly across community sizes (37% are in large cities, 13% in middle-sized communities, and 50% in small communities, overall). It is clear from Table 9, on the other hand, that secondary schools and special schools were more likely to serve large population areas. This suggests the need for a large population base to support these types of schools. Decreases in numbers of schools are occurring in both large cities and small communities, however. As Tables 10 and 11 show, there are relatively more students than schools in large cities (46% vs. 37%); that is, large schools tend to be in large cities. Otherwise, the distribution of nonpublic school students across community sizes mirrors the regularities seen in distribution of schools.

As shown in Tables 12 and 13, the average pupil/teacher ratio is not related to community size, nor are changes in this ratio between years. The pupil/teacher ratio decreased by roughly 1% per year in each of the three community-size categories. Per-pupil expenditures for current operations, on the other hand, were highest in large cities, as seen in Tables 14 and 15. However, secondary schools and special schools located in small communities had higher per-pupil expenditures than similar schools in other locations. This may explain the tendency for secondary and special schools to locate in large cities, noted above.

As shown in Exhibit C below, new schools were opening and old schools closing in all regions of the country. The largest percentages of new schools were in the Southeast and in the West, areas with greatest popula-

tion growth in recent years; however, these areas also saw the greatest percentage of closures of nonpublic schools during this period. These trends complement results obtained from the Current Population Survey (Bianchi, 1979), that private school enrollment increased between 1970 and 1978 in the West and South but decreased in the Northeast and North Central regions.

EXHIBIT C

Percentages of Schools Newly Opening and Newly Closing, by Region

REGION	New in 1977-78 / Open in 1977-78	New in 1978-79 / Open in 1978-79	Closed after 1976-77 / Open in 1976-77	Closed after 1977-78 / Open in 1977-78
Northeast (n=6196)	1.0%	0.3%	1.4%	2.8%
Southeast (n=4354)	1.7%	1.1%	2.1%	2.8%
Central (n=6350)	1.4%	0.8%	1.2%	2.3%
West (n=3604)	2.2%	1.0%	1.7%	3.4%
TOTAL (n=20504)	1.5%	0.7%	1.6%	2.8%

TABLE 8. NUMBER OF SCHOOLS BY RELIGIOUS AFFILIATION AND COMMUNITY SIZE

	COMMUNITY SIZE							
	> 250,000		50,000 to 250,000		< 50,000		TOTAL	
	Count	Percent Change	Count	Percent Change	Count	Percent Change	Count	Percent Change
RELIGIOUS AFFILIATION								
Catholic								
1976/1977	3789		1495		4682		9966	
1977/1978	3761	-1	1493	-0	4659	-0	9913	-1
1978/1979	3735	-1	1485	-1	4629	-1	9849	-1
Lutheran								
1976/1977	549		173		773		1495	
1977/1978	551	0	173	0	774	0	1498	0
1978/1979	545	-1	170	-2	770	-1	1485	-1
Other Religious Affiliation								
1976/1977	1436		488		2527		4451	
1977/1978	1449	1	492	1	2594	3	4535	2
1978/1979	1414	-2	494	0	2480	-4	4388	-3
No Religious Affiliation								
1976/1977	1735		455		1981		4171	
1977/1978	1726	-1	455	0	1946	-2	4127	-1
1978/1979	1641	-5	438	-4	1865	-4	3944	-4
TOTAL								
1976/1977	7509		2611		9963		20083	
1977/1978	7487	-0	2613	0	9973	0	20073	-0
1978/1979	7335	-2	2587	-1	9744	-2	19666	-2

TABLE 9. NUMBER OF SCHOOLS BY SCHOOL TYPE AND COMMUNITY SIZE

SCHOOL TYPE								
Elementary and Middle								
1976/1977	4686		1799		6480		12965	
1977/1978	4662	-1	1801	0	6471	-0	12934	-0
1978/1979	4592	-2	1796	-0	6361	-2	12749	-1
Secondary								
1976/1977	1116		335		1033		2484	
1977/1978	1109	-1	332	-1	1021	-1	2462	-1
1978/1979	1090	-2	328	-1	1000	-2	2418	-2
Combined Elementary and Secondary								
1976/1977	1107		351		1964		3422	
1977/1978	1119	1	353	1	1995	2	3467	1
1978/1979	1087	-3	344	-3	1920	-4	3351	-3
Other: Special, Vocational or Alternative								
1976/1977	600		126		486		1212	
1977/1978	597	-1	127	1	486	0	1210	-0
1978/1979	566	-5	119	-6	463	-5	1148	-5
TOTAL								
1976/1977	7509		2611		9963		20083	
1977/1978	7487	-0	2613	0	9973	0	20073	-0
1978/1979	7335	-2	2587	-1	9744	-2	19666	-2

TABLE 10. NUMBER OF STUDENTS(in thousands) BY RELIGIOUS AFFILIATION AND COMMUNITY SIZE

	COMMUNITY SIZE							
	> 250,000		50,000 to 250,000		< 50,000		TOTAL	
	Count	Percent Change	Count	Percent Change	Count	Percent Change	Count	Percent Change
RELIGIOUS AFFILIATION								
Catholic								
1976/1977	1570		539		1314		3423	
1977/1978	1533	-2	527	-2	1281	-2	3341	-2
1978/1979	1497	-2	518	-2	1255	-2	3270	-2
Lutheran								
1976/1977	95		28		89		212	
1977/1978	97	3	28	1	91	2	217	2
1978/1979	96	-1	28	1	93	1	217	0
Other Religious Affiliation								
1976/1977	361		91		321		773	
1977/1978	374	3	97	6	343	7	814	5
1978/1979	393	5	103	6	356	4	852	5
No Religious Affiliation								
1976/1977	349		67		341		758	
1977/1978	355	2	67	-0	346	1	768	1
1978/1979	341	-4	66	-2	340	-2	747	-3
TOTAL								
1976/1977	2376		726		2065		5167	
1977/1978	2359	-1	719	-1	2062	-0	5140	-1
1978/1979	2328	-1	715	-1	2043	-1	5086	-1

TABLE 11. NUMBER OF STUDENTS(in thousands) BY SCHOOL TYPE AND COMMUNITY SIZE

SCHOOL TYPE								
Elementary and Middle								
1976/1977	1309		479		1293		3081	
1977/1978	1286	-2	469	-2	1271	-2	3025	-2
1978/1979	1263	-2	465	-1	1261	-1	2989	-1
Secondary								
1976/1977	597		150		333		1080	
1977/1978	589	-1	149	-1	326	-2	1064	-1
1978/1979	590	0	149	0	329	1	1069	0
Combined Elementary and Secondary								
1976/1977	411		88		406		905	
1977/1978	430	5	93	6	430	6	952	5
1978/1979	424	-1	92	-0	421	-2	938	-1
Other: Special, Vocational or Alternative								
1976/1977	58		9		34		101	
1977/1978	54	-8	9	-3	36	6	98	-3
1978/1979	50	-7	8	-6	32	-9	90	-8
TOTAL								
1976/1977	2376		726		2065		5167	
1977/1978	2359	-1	719	-1	2062	-0	5140	-1
1978/1979	2328	-1	715	-1	2043	-1	5086	-1

TABLE 12. PUPIL/TEACHER RATIO BY RELIGIOUS AFFILIATION AND COMMUNITY SIZE

	COMMUNITY SIZE							
	> 250,000		50,000 to 250,000		< 50,000		TOTAL	
	Average	Percent Change	Average	Percent Change	Average	Percent Change	Average	Percent Change
RELIGIOUS AFFILIATION								
Catholic								
1976/1977	23.7		23.8		22.9		23.4	
1977/1978	23.2	-2	23.2	-2	22.4	-2	22.9	-2
1978/1979	23.0	-1	23.1	-0	22.5	0	22.8	-0
Lutheran								
1976/1977	22.0		23.5		21.3		21.9	
1977/1978	21.3	-3	22.3	-5	20.7	-3	21.2	-3
1978/1979	21.3	0	22.3	0	21.1	2	21.3	1
Other Religious Affiliation								
1976/1977	15.4		15.6		15.0		15.3	
1977/1978	15.1	-2	15.1	-3	14.4	-4	14.8	-3
1978/1979	15.4	2	15.7	4	14.9	4	15.2	3
No Religious Affiliation								
1976/1977	12.0		11.9		12.4		12.2	
1977/1978	11.3	-6	11.1	-6	11.8	-5	11.5	-6
1978/1979	11.5	2	11.5	3	12.2	4	11.8	3
TOTAL								
1976/1977	19.3		20.5		18.7		19.2	
1977/1978	18.6	-4	19.8	-4	18.0	-4	18.5	-4
1978/1979	18.6	0	19.9	1	18.2	2	18.6	1

TABLE 13. PUPIL/TEACHER RATIO BY SCHOOL TYPE AND COMMUNITY SIZE

SCHOOL TYPE								
Elementary and Middle								
1976/1977	23.4		24.0		22.5		23.1	
1977/1978	22.8	-3	23.3	-3	21.9	-3	22.5	-3
1978/1979	22.6	-1	23.0	-1	22.1	1	22.5	-0
Secondary								
1976/1977	18.8		18.0		15.1		17.4	
1977/1978	18.4	-2	17.4	-3	14.7	-3	16.9	-3
1978/1979	18.4	0	17.8	2	14.9	1	17.1	1
Combined Elementary and Secondary								
1976/1977	14.3		14.9		15.3		14.8	
1977/1978	13.8	-3	14.5	-3	14.8	-4	14.3	-3
1978/1979	14.0	2	15.0	4	15.0	1	14.5	2
Other: Special, Vocational or Alternative								
1976/1977	8.8		7.8		7.5		8.2	
1977/1978	7.3	-17	6.6	-15	6.6	-11	7.0	-15
1978/1979	7.4	1	6.8	2	7.0	6	7.2	3
TOTAL								
1976/1977	19.3		20.5		18.7		19.2	
1977/1978	18.6	-4	19.8	-4	18.0	-4	18.5	-4
1978/1979	18.6	0	19.9	1	18.2	2	18.6	1

TABLE 14. PER-PUPIL EXPENDITURES BY RELIGIOUS AFFILIATION AND COMMUNITY SIZE

	COMMUNITY SIZE							
	> 250,000		50,000 to 250,000		< 50,000		TOTAL	
	Average	Percent Change	Average	Percent Change	Average	Percent Change	Average	Percent Change
RELIGIOUS AFFILIATION								
Catholic								
1976/1977	$613		$551		$546		[illegible]78	
1977/1978	$656	7	$593	8	$577	6	$616	7
1978/1979								
Lutheran								
1976/1977	$776		$640		$768		$755	
1977/1978	$780	0	$648	1	$734	-4	$744	-1
1978/1979								
Other Religious Affiliation								
1976/1977	$1,383		$2,000		$1,351		$1,442	
1977/1978	$1,314	-5	$1,961	-2	$1,282	-5	$1,377	-5
1978/1979								
No Religious Affiliation								
1976/1977	$2,525		$2,282		$2,383		$2,440	
1977/1978	$2,546	1	$2,315	1	$2,354	-1	$2,439	-0
1978/1979								
TOTAL								
1976/1977	$1,018		$897		$985		$988	
1977/1978	$1,050	3	$940	5	$1,000	2	$1,014	3
1978/1979								

TABLE 15. PER-PUPIL EXPENDITURES BY SCHOOL TYPE AND COMMUNITY SIZE

SCHOOL TYPE								
Elementary and Middle								
1976/1977	$626		$681		$577		$614	
1977/1978	$655	5	$736	8	$603	5	$646	5
1978/1979								
Secondary								
1976/1977	$1,050		$1,069		$1,686		$1,248	
1977/1978	$1,103	5	$1,135	6	$1,748	4	$1,305	5
1978/1979								
Combined Elementary and Secondary								
1976/1977	$1,819		$1,398		$1,291		$1,541	
1977/1978	$1,774	-2	$1,289	-8	$1,219	-6	$1,475	-4
1978/1979								
Other: Special, Vocational or Alternative								
1976/1977	$3,825		$4,731		$6,005		$4,634	
1977/1978	$4,125	8	$4,949	5	$5,676	-5	$4,762	3
1978/1979								
TOTAL								
1976/1977	$1,018		$897		$985		$988	
1977/1978	$1,050	3	$940	5	$1,000	2	$1,014	3
1978/1979								

Sex of students served. In the next set of tables, the extent to which nonpublic schools are coeducational is displayed. In the Educational Amendments of 1972, Congress specifically included private schools in Title IX, prohibition of sex discrimination, although it exempted schools affiliated with religious organizations whose tenets the Title would violate. Discrimination in admissions, moreover, was covered only for postsecondary institutions. Thus, nonpublic schools for boys only or for girls only are not in violation of the law. Nevertheless, based on Tables 15 and 17, only about 7% of the nonpublic schools serve a single sex, split equally between schools for boys only and schools for girls only. These include both religiously affiliated and non-affiliated schools. Over a third of the secondary schools, however, are for either boys only or girls only. Because secondary schools serving a single sex are larger than other secondary schools, as shown in Table 19, 40% of nonpublic secondary school students are in schools serving only young people of their own sex.

Pupil/teacher ratios are smaller in single sex schools, especially in schools for girls only, as shown in Tables 20 and 21. At the secondary level, however, the differences disappear. Per-pupil expenditures, on the other hand, are highest in schools for boys only, as shown in Tables 22 and 23. This difference occurs across all categories of religious affiliation and school type, and it indicates either that a large portion of the expenditures for boys' schools are not for classroom teachers or that teachers at schools for boys only are paid more than teachers at schools for girls only. Both hypotheses are somewhat supported. First, there was a correlation of .25 between per-pupil expenditures and percentage of classroom teachers who are male. On the other hand, there were on the average 18 staff other than classroom teachers in boys-only schools, 14 in girls-only schools, and 7 in schools serving both sexes, in 1977-78.

TABLE 16. NUMBER OF SCHOOLS BY RELIGIOUS AFFILIATION AND SEX OF STUDENTS SERVED

	SEX OF STUDENTS SERVED							
	Boys Only		Girls Only		Both Boys and Girls		TOTAL	
	Count	Percent Change	Count	Percent Change	Count	Percent Change	Count	Percent Change
RELIGIOUS AFFILIATION								
Catholic								
1976/1977	329		477		9160		9966	
1977/1978	363	10	513	8	9037	-1	9913	-1
1978/1979	363	0	477	-7	9009	-0	9849	-1
Lutheran								
1976/1977	4		1		1490		1495	
1977/1978	7	75	1	0	1490	0	1498	0
1978/1979	5	-29			1480	-1	1485	-1
Other Religious Affiliation								
1976/1977	121		76		4254		4451	
1977/1978	163	35	86	13	4286	1	4535	2
1978/1979	146	-10	83	-3	4159	-3	4388	-3
No Religious Affiliation								
1976/1977	190		87		3894		4171	
1977/1978	204	7	82	-6	3841	-1	4127	-1
1978/1979	186	-9	79	-4	3679	-4	3944	-4
TOTAL								
1976/1977	644		641		18798		20083	
1977/1978	737	14	682	6	18654	-1	20073	-0
1978/1979	700	-5	639	-6	18327	-2	19666	-2

TABLE 17. NUMBER OF SCHOOLS BY SCHOOL TYPE AND SEX OF STUDENTS SERVED

SCHOOL TYPE								
Elementary and Middle								
1976/1977	78		51		12836		12965	
1977/1978	119	53	67	31	12748	-1	12934	-0
1978/1979	107	-10	57	-15	12585	-1	12749	-1
Secondary								
1976/1977	392		485		1607		2484	
1977/1978	420	7	510	5	1532	-5	2462	-1
1978/1979	404	-4	478	-6	1536	0	2418	-2
Combined Elementary and Secondary								
1976/1977	84		68		3270		3422	
1977/1978	96	14	69	1	3302	1	3467	
1978/1979	95	-1	73	6	3183	-4	3351	-3
Other: Special, Vocational or Alternative								
1976/1977	90		37		1085		1212	
1977/1978	102	13	36	-3	1072	-1	1210	-0
1978/1979	94	-3	31	-14	1023	-5	1148	-5
TOTAL								
1976/1977	644		641		18798		20083	
1977/1978	737	14	682	6	18654	-1	20073	-0
1978/1979	700	-5	639	-6	18327	-2	19666	-2

TABLE 18. NUMBER OF STUDENTS(in thousands) BY RELIGIOUS AFFILIATION AND SEX OF STUDENTS SERVED

	SEX OF STUDENTS SERVED							
	Boys Only		Girls Only		Both Boys and Girls		TOTAL	
	Court	Percent Change	Count	Percent Change	unt	Percent Change	Count	Percent Change
RELIGIOUS AFFILIATION								
Catholic								
1976/1977	204		206		3013		3423	
1977/1978	218	7	219	6	2904	-4	3341	-2
1978/1979	217	-0	209	-4	2843	-2	3270	-2
Lutheran								
1976/1977					212		212	
1977/1978	*		*		216	2	217	2
1978/1979					217	0	217	0
Other Religious Affiliation								
1976/1977	24		20		729		773	
1977/1978	33	39	22	11	758	4	814	5
1978/1979	34	0	23	4	795	5	852	5
No Religious Affiliation								
1976/1977	36		20		702		758	
1977/1978	37	5	20	-0	711	1	768	1
1978/1979	35	-5	21	5	690	-3	747	-3
TOTAL								
1976/1977	264		246		4657		5167	
1977/1978	290	10	261	6	4589	-1	5140	-1
1978/1979	287	-1	253	-3	4544	-1	5086	-1

TABLE 19. NUMBER OF STUDENTS(in thousands) BY SCHOOL TYPE AND SEX OF STUDENTS SERVED

SCHOOL TYPE								
Elementary and Middle								
1976/1977	12		14		3055		3081	
1977/1978	21	80	17	26	2987	-2	3025	-2
1978/1979	20	-5	17	-5	2953	-1	2989	-1
Secondary								
1976/1977	214		204		662		1080	
1977/1978	225	5	213	5	626	-6	1064	-1
1978/1979	224	-1	205	-4	640	2	1069	0
Combined Elementary and Secondary								
1976/1977	32		27		847		905	
1977/1978	37	15	29	8	887	5	952	5
1978/1979	36	-1	30	7	871	-2	938	-1
Other: Special, Vocational or Alternative								
1976/1977	6		2		93		101	
1977/1978	7	15	2	-9	89	-4	98	-3
1978/1979	6	-4	1	-28	82	-7	90	-8
TOTAL								
1976/1977	264		246		4657		5167	
1977/1978	290	10	261	6	4589	-1	5140	-1
1978/1979	287	-1	253	-3	4546	-1	5086	-1

*Fewer than five schools in this category for at least one year.

TABLE 20. PUPIL/TEACHER RATIO BY RELIGIOUS AFFILIATION AND SEX OF STUDENTS SERVED

	SEX OF STUDENTS SERVED							
	Boys Only		Girls Only		Both Boys and Girls		TOTAL	
	Average	Percent Change	Average	Percent Change	Average	Percent Change	Average	Percent Change
RELIGIOUS AFFILIATION								
Catholic								
1976/1977	19.9		17.1		24.3		23.4	
1977/1978	19.4	-2	16.9	-1	23.9	-2	22.9	-2
1978/1979	19.4	-0	17.0	1	23.7	-1	22.8	-0
Lutheran								
1976/1977	*		*		21.9		21.9	
1977/1978					21.2	-3	21.2	-3
1978/1979					21.3	1	21.3	1
Other Religious Affiliation								
1976/1977	13.0		13.0		15.4		15.3	
1977/1978	13.4	3	12.7	-2	14.9	-3	14.3	-3
1978/1979	13.4	1	12.3	-3	15.4	3	15.2	3
No Religious Affiliation								
1976/1977	10.5		9.8		12.3		12.2	
1977/1978	9.9	-6	9.1	-7	11.6	-6	11.5	-6
1978/1979	10.2	3	9.7	6	12.0	3	11.8	3
TOTAL								
1976/1977	17.8		15.7		19.6		19.2	
1977/1978	16.5	-3	15.5	-2	18.8	-4	18.5	-4
1978/1979	16.7	1	15.5	0	19.0	1	18.6	1

TABLE 21. PUPIL/TEACHER RATIO BY SCHOOL TYPE AND SEX OF STUDENTS SERVED

SCHOOL TYPE								
Elementary and Middle								
1976/1977	13.6		17.1		23.2		23.1	
1977/1978	14.6	8	17.1	-0	22.6	-3	22.5	-3
1978/1979	15.1	3	15.1	-11	22.6	-0	22.5	-0
Secondary								
1976/1977	18.5		16.7		17.3		17.4	
1977/1978	18.1	-2	16.5	-2	16.7	-3	16.9	-3
1978/1979	18.1	0	16.6	1	16.9	1	17.1	1
Combined Elementary and Secondary								
1976/1977	13.5		11.4		15.0		14.8	
1977/1978	13.5	-0	10.7	-6	14.5	-3	14.3	-3
1978/1979	13.7	2	11.4	6	14.7	2	14.5	2
Other: Special, Vocational or Alternative								
1976/1977	8.5		6.9		8.2		8.2	
1977/1978	7.3	-14	7.2	5	6.9	-16	7.0	-15
1978/1979	7.4	0	6.3	-13	7.2	3	7.2	3
TOTAL								
1976/1977	17.0		15.7		19.6		19.2	
1977/1978	16.5	-3	15.5	-2	18.8	-4	18.5	-4
1978/1979	16.7	1	15.5	0	19.0	1	18.6	1

*Fewer than five schools in this category for at least one year.

TABLE 22. PER-PUPIL EXPENDITURES BY RELIGIOUS AFFILIATION AND SEX OF STUDENTS SERVED

	SEX OF STUDENTS SERVED							
	Boys Only		Girls Only		Both Boys and Girls		TOTAL	
	Average	Percent Change	Average	Percent Change	Average	Percent Change	Average	Percent Change
RELIGIOUS AFFILIATION								
Catholic								
1976/1977	$1,074		$887		$523		$578	
1977/1978	$1,127	5	$902	2	$556	6	$616	7
1978/1979								
Lutheran								
1976/1977	*		*		$752		$755	
1977/1978					$743	-1	$744	-1
1978/1979								
Other Religious Affiliation								
1976/1977	$2,233		$1,695		$1,409		$1,442	
1977/1978	$2,177	-2	$1,627	-4	$1,334	-5	$1,377	-5
1978/1979								
No Religious Affiliation								
1976/1977	$3,586		$3,183		$2,360		$2,440	
1977/1978	$3,637	1	$3,246	2	$2,353	-0	$2,439	-0
1978/1979								
TOTAL								
1976/1977	$1,521		$1,140		$949		$988	
1977/1978	$1,573	3	$1,144	0	$972	2	$1,014	3
1978/1979								

TABLE 23. PER-PUPIL EXPENDITURES BY SCHOOL TYPE AND SEX OF STUDENTS SERVED

SCHOOL TYPE								
Elementary and Middle								
1976/1977	$1,812		$917		$608		$614	
1977/1978	$1,627	-10	$839	-8	$638	5	$646	5
1978/1979								
Secondary								
1976/1977	$1,316		$1,009		$1,301		$1,248	
1977/1978	$1,363	4	$1,015	1	$1,383	6	$1,305	5
1978/1979								
Combined Elementary and Secondary								
1976/1977	$2,089		$2,031		$1,505		$1,541	
1977/1978	$2,129	2	$2,092	3	$1,429	-5	$1,476	-4
1978/1979								
Other: Special, Vocational or Alternative								
1976/1977	$5,312		$4,198		$4,601		$4,634	
1977/1978	$5,354	1	$4,175	-1	$4,730	3	$4,762	3
1978/1979								
TOTAL								
1976/1977	$1,521		$1,140		$949		$988	
1977/1978	$1,573	3	$1,144	0	$972	2	$1,014	3
1978/1979								

*Fewer than five schools in this category for at least one year.

Day/boarding facilities. Differences between day and boarding schools are displayed in Tables 24 through 31. These differences are of interest for two reasons. First, day schools draw students only from nearby areas whereas boarding schools can have a wider drawing area. Thus, day schools bear a more direct relationship to the nearby public schools. Second, boarding schools must in general draw students from more affluent families, although exceptions occur. In most boarding schools, parents are called on to pay not only tuition but also boarding costs to send their children to these schools.

Boarding schools are most frequently not religiously affiliated, as seen in Table 24, although many religiously affiliated schools, other than Catholic and Lutheran, accept both day and boarding students. Among the few Catholic and Lutheran boarding schools, a substantial percentage were closing or converting to day schools between 1976 and 1978. Most boarding schools, as shown in Table 25, were either secondary schools or special, vocational, and alternative schools. These results are corroborated by student counts in Tables 26 and 27. Comparison among these four tables indicates, moreover, that boarding schools are substantially smaller than day schools, except for special, vocational, and alternative schools.

Pupil/teacher ratios, as shown in Tables 28 and 29, are half in boarding schools what they are in day schools, while ratios in day schools differ little from ratios in public schools. Day schools with no religious affiliation or of special types are exceptions to this, having pupil/teacher ratios more nearly like boarding schools. Similarly, as shown in Tables 30 and 31, per-pupil expenditures for current operations in schools with boarding facilities are two to three times those of day schools of the same type, even though these expenditures exclude specific boarding expenditures. Moreover, these differences widened from 1976-77 to 1977-78, as increases in expenditures in day schools were below the national inflation rate but increases in expenditures in boarding (only) schools increased at 17%, twice the national inflation rate.

TABLE 24. NUMBER OF SCHOOLS BY RELIGIOUS AFFILIATION AND DAY-BOARDING MIX

	DAY-BOARDING MIX							
	Day Schools		Boarding Schools		Both Day and Boarding		TOTAL	
	Count	Percent Change	Count	Percent Change	Count	Percent Change	Count	Percent Change
RELIGIOUS AFFILIATION								
Catholic								
1976/1977	9686		87		193		9966	
1977/1978	9648	-0	71	-18	194	1	9913	-1
1978/1979	9618	-0	60	-15	171	-12	9849	-1
Lutheran								
1976/1977	1456		12		27		1495	
1977/1978	1468	1	7	-42	23	-15	1498	0
1978/1979	1456	-1	7	0	22	-4	1485	-1
Other Religious Affiliation								
1976/1977	4096		84		271		4451	
1977/1978	4170	2	74	-12	291	7	4535	2
1978/1979	4026	-3	67	-9	295	1	4388	-3
No Religious Affiliation								
1976/1977	3514		216		441		4171	
1977/1978	3417	-3	218	1	492	12	4127	-1
1978/1979	3222	-6	254	17	468	-5	3944	-4
TOTAL								
1976/1977	18752		399		932		20083	
1977/1978	18703	-0	370	-7	1000	7	20073	-0
1978/1979	18322	-2	388	5	956	-4	19666	-2

TABLE 25. NUMBER OF SCHOOLS BY SCHOOL TYPE AND DAY-BOARDING MIX

SCHOOL TYPE								
Elementary and Middle								
1976/1977	12774		75		116		12965	
1977/1978	12791	0	46	-39	97	-16	12934	-0
1978/1979	12614	-1	38	-17	97	0	12749	-1
Secondary								
1976/1977	1970		107		407		2484	
1977/1978	1918	-3	111	4	433	6	2462	-1
1978/1979	1881	-2	112	1	425	-2	2418	-2
Combined Elementary and Secondary								
1976/1977	3192		48		182		3422	
1977/1978	3209	1	45	-6	213	17	3467	1
1978/1979	3108	-3	48	7	195	-8	3351	-3
Other: Special, Vocational or Alternative								
1976/1977	816		169		227		1212	
1977/1978	785	-4	168	-1	257	13	1210	-0
1978/1979	719	-8	190	13	239	-7	1148	-5
TOTAL								
1976/1977	18752		399		932		20083	
1977/1978	18703	-0	370	-7	1000	7	20073	-0
1978/1979	18322	-2	388	5	956	-4	19666	-2

TABLE 26. NUMBER OF STUDENTS(in thousands) BY RELIGIOUS AFFILIATION AND DAY-BOARDING MIX

	DAY-BOARDING MIX							
	Day Schools		Boarding Schools		Both Day and Boarding		TOTAL	
	Count	Percent Change	Count	Percent Change	Count	Percent Change	Count	Percent Change
RELIGIOUS AFFILIATION								
Catholic								
1976/1977	3367		14		42		3423	
1977/1978	3292	-2	8	-45	41	-3	3341	-2
1978/1979	3225	-2	6	-22	38	-7	3270	-2
Lutheran								
1976/1977	209		0		3		212	
1977/1978	213	2	0	-17	4	11	217	2
1978/1979	213	0	0	49	4	-4	217	0
Other Religious Affiliation								
1976/1977	708		9		57		773	
1977/1978	744	5	8	-4	61	8	814	5
1978/1979	772	4	9	14	71	15	852	5
No Religious Affiliation								
1976/1977	650		23		86		758	
1977/1978	651	0	23	2	94	9	768	1
1978/1979	632	-3	27	17	88	-6	747	-3
TOTAL								
1976/1977	4933		46		188		5167	
1977/1978	4900	-1	39	-14	200	6	5140	-1
1978/1979	4842	-1	43	9	200	0	5086	-1

TABLE 27. NUMBER OF STUDENTS(in thousands) BY SCHOOL TYPE AND DAY-BOARDING MIX

SCHOOL TYPE								
Elementary and Middle								
1976/1977	3051		10		20		3081	
1977/1978	3005	-2	5	-52	16	-21	3025	-2
1978/1979	2964	-1	5	3	20	31	2989	-1
Secondary								
1976/1977	977		16		87		1080	
1977/1978	956	-2	15	-4	93	6	1064	-1
1978/1979	954	-0	17	8	98	5	1069	0
Combined Elementary and Secondary								
1976/1977	838		8		59		905	
1977/1978	878	5	8	-5	66	12	952	5
1978/1979	870	-1	8	6	60	-9	938	-1
Other: Special, Vocational or Alternative								
1976/1977	66		12		23		101	
1977/1978	60	-9	11	-3	26	14	98	-3
1978/1979	55	-9	13	14	23	-13	90	-8
TOTAL								
1976/1977	4933		46		188		5167	
1977/1978	4900	-1	39	-14	200	6	5140	-1
1978/1979	4842	-1	43	9	200	0	5086	-1

TABLE 28. PUPIL/TEACHER RATIO BY RELIGIOUS AFFILIATION AND DAY-BOARDING MIX

	DAY-BOARDING MIX							
	Day Schools		Boarding Schools		Both Day and Boarding		TOTAL	
	Average	Percent Change	Average	Percent Change	Average	Percent Change	Average	Percent Change
RELIGIOUS AFFILIATION								
Catholic								
1976/1977	23.7		12.7		12.9		23.4	
1977/1978	23.3	-2	9.3	-27	10.5	-19	22.9	-2
1978/1979	23.2	-1	7.9	-14	11.8	12	22.8	-0
Lutheran								
1976/1977	22.1		12.6		14.5		21.9	
1977/1978	21.3	-4	9.0	-28	15.3	6	21.2	-3
1978/1979	21.5	1	12.2	35	15.7	3	21.3	1
Other Religious Affiliation								
1976/1977	15.8		10.3		11.6		15.3	
1977/1978	15.3	-3	8.9	-13	11.0	-5	14.8	-3
1978/1979	15.7	2	9.7	9	12.2	11	15.2	3
No Religious Affiliation								
1976/1977	12.8		10.0		9.3		12.2	
1977/1978	12.2	-4	8.6	-13	8.5	-9	11.5	-6
1978/1979	12.6	3	9.5	10	8.5	-0	11.8	3
TOTAL								
1976/1977	20.0		10.8		10.7		19.2	
1977/1978	19.4	-3	8.8	-18	9.6	-10	18.5	-4
1978/1979	19.5	1	9.3	6	10.2	6	18.6	1

TABLE 29. PUPIL/TEACHER RATIO BY SCHOOL TYPE AND DAY-BOARDING MIX

SCHOOL TYPE								
Elementary and Middle								
1976/1977	23.2		16.9		15.0		23.1	
1977/1978	22.7	-2	12.2	-23	9.7	-35	22.5	-3
1978/1979	22.6	-0	13.0	6	16.1	65	22.5	-0
Secondary								
1976/1977	18.6		10.4		10.7		17.4	
1977/1978	18.4	-2	9.4	-10	10.1	-5	16.9	-3
1978/1979	18.4	1	9.9	5	10.6	5	17.1	1
Combined Elementary and Secondary								
1976/1977	15.1		11.6		11.9		14.8	
1977/1978	14.6	-3	10.5	-9	11.2	-6	14.3	-3
1978/1979	14.9	2	10.9	3	11.3	1	14.5	2
Other: Special, Vocational or Alternative								
1976/1977	8.7		8.2		7.1		8.2	
1977/1978	7.4	-15	6.7	-18	6.2	-13	7.0	-15
1978/1979	7.9	6	7.3	10	5.8	-6	7.2	3
TOTAL								
1976/1977	20.0		10.8		10.7		19.2	
1977/1978	19.4	-3	8.8	-18	9.6	-10	18.5	-4
1978/1979	19.5	1	9.3	6	10.2	6	18.6	1

TABLE 30. PER-PUPIL EXPENDITURES BY RELIGIOUS AFFILIATION AND DAY-BOARDING MIX

	DAY-BOARDING MIX							
	Day Schools		Boarding Schools		Both Day and Boardin		TOTAL	
	Average	Percent Change	Average	Percent Change	Average	Percent Change	Average	Percent Change
RELIGIOUS AFFILIATION								
Catholic								
1976/1977	$555		$2,029		$1,859		$578	
1977/1978	$594	7	$3,206	58	$1,888	2	$616	7
1978/1979								
Lutheran								
1976/1977	$728		$2,273		$2,288		$755	
1977/1978	$725	-0	$3,233	42	$1,607	-30	$744	-1
1978/1979								
Other Religious Affiliation								
1976/1977	$1,294		$4,174		$2,871		$1,442	
1977/1978	$1,225	-5	$4,274	2	$2,837	-1	$1,377	-5
1978/1979								
No Religious Affiliation								
1976/1977	$2,035		$4,829		$4,861		$2,440	
1977/1978	$2,009	-1	$4,992	3	$4,788	-2	$2,439	-0
1978/1979								
TOTAL								
1976/1977	$864		$3,809		$3,545		$988	
1977/1978	$884	2	$4,470	17	$3,539	-0	$1,014	3
1978/1979								

TABLE 31. PER-PUPIL EXPENDITURES BY SCHOOL TYPE AND DAY-BOARDING MIX

SCHOOL TYPE								
Elementary and Middle								
1976/1977	$606		$1,233		$1,604		$614	
1977/1978	$638	5	$2,422	96	$1,664	4	$646	5
1978/1979								
Secondary								
1976/1977	$988		$3,996		$3,665		$1,248	
1977/1978	$1,036	5	$4,036	1	$3,626	-1	$1,305	5
1978/1979								
Combined Elementary and Secondary								
1976/1977	$1,452		$3,137		$2,589		$1,541	
1977/1978	$1,384	-5	$3,179	1	$2,499	-3	$1,476	-4
1978/1979								
Other: Special, Vocational or Alternative								
1976/1977	$3,463		$6,235		$7,221		$4,634	
1977/1978	$3,421	-1	$6,821	9	$6,976	-3	$4,762	3
1978/1979								
TOTAL								
1976/1977	$864		$3,809		$3,545		$988	
1977/1978	$884	2	$4,470	17	$3,539	-0	$1,014	3
1978/1979								

Federal program participation. Finally, as apparent in Tables 32 and 33, nonpublic schools are participating in federal assistance programs in large numbers. Nearly all Catholic schools, along with nearly one-half of other nonpublic schools, participate in Title IV-B programs, including library resources, instructional equipment, testing, counseling, and guidance services. One-half of the Catholic schools and one-seventh of other nonpublic schools participate in Title I programs of compensatory education for educationally disadvantaged children. Nearly all nonpublic schools participating in Title I served elementary grades, a characteristic similar to public schools. Nearly three in four Catholic schools and over one-third of other nonpublic schools participate in school lunch or special milk programs. Finally, there was a noticeable decrease in reported Title IV-C participation between 1977-78 and 1978-79. Title IV-C supports supplementary educational centers and services, exemplary programs, dropout prevention, and health and nutrition services. The decreased frequency might reflect the replacement of "health and nutrition services" by "etc." in the 1978-79 form, because it is not unreasonable to suppose that respondents were not conversant with program numbers and had responded positively to this item in earlier years because of other health or nutrition assistance (e.g., school lunch support).

Table 32. Number of Schools Receiving Federal Assistance, by Religious Affiliation

	Catholic		Lutheran		Other religious affiliation		No affiliation		Total	
	Number	Percent Change	Number	Percent Change	Number	Percent Change	Number	Percent Change	Number	Percent Change
Title I										
1976-77	4725		268		617		566		6176	
1977-78	4857	3	319	19	632	2	596	5	6404	4
1978-79	4974	2	300	-6	596	-6	576	-3	6446	1
Title IVB										
1976-77	9129		869		1594		1744		13336	
1977-78	9207	1	894	3	1620	2	1769	1	13490	1
1978-79	9172	-0	905	1	1599	-1	1728	-2	13404	-1
Title IVC										
1976-77	1420		142		125		241		1928	
1977-78	1636	15	151	6	135	8	261	8	2183	13
1978-79	1242	-24	104	-31	123	-10	229	-12	1698	-22
School Lunch										
1976-77	6996		992		1153		1532		10673	
1977-78	7186	3	1072	8	1166	1	1535	0	10959	3
1978-79	7074	-2	1046	-2	1139	-2	1488	-3	10747	-2
Other										
1976-77	1163		65		403		295		1926	
1977-78	1062	-9	49	-25	389	-3	322	9	1822	-5
1978-79	884	-17	30	-39	342	-12	317	-1	1573	-14

Table 33. Number of Schools Receiving Federal Assistance, By School Type

	Elem./ Middle		Combined Elem./Sec.		Secondary		Other		Total	
Title I										
1976-77	5103		412		283		378		6176	
1977-78	5302	4	429	4	280	-1	393	4	6404	4
1978-79	5387	2	410	-4	285	2	364	-7	6446	1
Title IVB										
1976-77	9669		1366		1847		454		13336	
1977-78	9759	1	1395	2	1856	1	480	6	13490	1
1978-79	9707	-1	1359	-3	1870	1	468	-2	13404	-1
Title IVC										
1976-77	1485		170		226		47		1928	
1977-78	1701	15	193	14	228	1	61	30	2183	13
1978-79	1292	-24	162	-16	192	-16	52	-15	1698	-22
School Lunch										
1976-77	8060		1083		1057		473		10673	
1977-78	8298	3	1104	2	1073	2	484	2	10959	3
1978-79	8152	-2	1071	-3	1058	-2	466	-4	10747	-2
Other										
1976-77	1264		205		233		224		1926	
1977-78	1193	-6	219	7	192	-18	218	-3	1822	-5
1978-79	1008	-16	202	-8	158	-18	205	-6	1573	-14

References

Bianchi, S. M. Trends in private school enrollment in the 1970's. Washington, DC: U.S. Bureau of the Census, Population Division, Education and Social Stratification Branch, 1979.

Bredeweg, F. H. Catholic elementary schools and their finances. Washington, DC: National Catholic Education Association, 1979(a).

Bredeweg, F. H. Catholic high schools and their finances. Washington, DC: National Catholic Education Association, 1979(b).

McLaughlin, D. H. (ed.). Meaning and significance in the analysis of the nonpublic school survey. SAGE Technical Report No. 2. Palo Alto, CA: American Institutes for Research, 1979. (Prepared under Contract No. 300-78-0150 with the National Center for Education Statistics.)

McLaughlin, D. H., and Wise, L. L. Report on the imputation of missing data on the 1977-78 nonpublic school survey. Palo Alto, CA: American Institutes for Research, Statistical Analysis Group in Education (SAGE), 1979. (Prepared under Contract No. 300-78-0150 with the National Center for Education Statistics.)

APPENDIX A

Future Survey Plans

Overall Strategy

It is essential that any government sponsored data collection involving a burden on nonpublic school staff be a cooperative effort serving the goals both of nonpublic school administrators and public policy planners. Especially because of the recent threats to nonpublic schools by the Internal Revenue Service, nonpublic school administrators cannot be blamed if they seriously question requests for information from the federal government. Although the subsequent congressional action to remove, or at least delay, the IRS threat may improve relations, nonpublic school administrators are aware that the information they provide to a federal survey may form the basis for policies inimical to their well-being. Therefore, no effort should be undertaken that does not involve prior participation, with expectation of endorsement, by representatives of the nonpublic school sector.

This situation implies the need for an integrated federal data collection and policy development plan, because the path that policy developers take will affect the willingness of nonpublic schools to provide data, and the availability of data will play a significant role in determining the match of federal policies to the needs of nonpublic schools. This plan should take into account the data needs of all parties, including state and local public education agencies, in order to minimize the redundancy and maximize the comparability of data collected. Therefore, the overall strategy should include a meeting or series of meetings, along with ongoing communication links, in order to prepare a workable comprehensive plan.

As a step in the direction of a comprehensive plan, SAGE has carried out a series of analyses of the three years of data collected by NCES in order to identify and evaluate data needs. The focus of these analyses concerned the extent of change in variable values from year to year. For the purposes of reporting aggregate statistics, the frequency with which data need to be collected is directly related to

the rate of change expected. To provide a context for these analyses, the following sections (1) summarize issues and data needs and (2) present criteria for evaluating alternative data collection designs. On the basis of the knowledge available to SAGE as of October 1979* and the analyses reported here, recommendations were made for a particular data collection strategy. These recommendations are presented as a starting point for discussion, to be refined as new knowledge is contributed by other parties.

Issues and Data Needs

The major issues confronting educational policy planners can be broken down into four categories: financial, characteristics of students served, quality of instruction, and organizational (Russ-Eft, Rubin, & Holman, 1979). The financial issues are two-faceted, concerning both the effects of nonpublic schools on public school finances and the effects of economic pressures on the viability of nonpublic schools. A surge of enrollments in nonpublic schools, combined with the declining birth rate, could produce serious problems for local public schools as fewer voters have a direct interest in supporting the schools. On the other hand, because 11% of the nation's elementary and secondary students are in nonpublic schools, the demise of a substantial portion of nonpublic schools could put substantial additional pressures on public schools. Policy decisions concerning tuition tax credits, for example, must address these issues.

Issues about student characteristics center on three concerns. Foremost, from the federal perspective of trying to achieve equal opportunity, is the concern over alleged growth of private schools that promote racial segregation. The recent IRS proposal to remove the tax exempt status of nonpublic schools that have been shown to be racially discriminatory and the congressional backlash it evoked are indicative of the importance of this issue. Second, concern exists over who will benefit from

* Changes in the nonpublic school survey data file introduced by reimputing missing data after October 1979 are not reflected in Appendix A. These changes would not affect the conclusions drawn, however.

federal aid to nonpublic schools--what percentage of the parents of nonpublic school children are affluent? How much increase in private or parochial school enrollment could be expected by introducing tuition tax credits? Are the children of affluent parents concentrated in certain types of schools? Third, there is concern that nonpublic schools will attract the best students out of public schools and avoid the problem students. A recent court ruling in New Jersey held that a private school could be excluded by surrounding public schools from an athletic league because, unlike the public schools, it could recruit students. This concern has important implications for attempts to evaluate the quality of instruction also, because variation in test scores of graduates is a function as much of student characteristics as of school characteristics.

Issues about quality of instruction center on whether children in some nonpublic schools receive the education necessary for participating in society as adults. While there is little question that the majority of nonpublic schools provide high quality instruction, that may not be true for all of them. On the other side of the coin, the emergence of the "voucher" issue, coupled with the alleged declining productivity and rising costs of public schools, will lead to greater needs for evaluation of the effectiveness of nonpublic and public schools alike.

Finally, there are the issues we have labeled "organizational." These issues center at present on the relations of public policy to nonpublic schools. Do public school districts have the responsibility to bus children attending nonpublic schools? Should nonpublic school students receive benefits of special federal programs on the premises of nonpublic schools or be shipped to public schools for those programs? How can public policies for parochial school students be developed that are not in conflict with the constitutional separation of church and state? These issues are manifestations of the overall questioning of the public role and responsibilities in education.

These issues imply a set of related data needs. For each of these data needs, questions about the precision and frequency required must

be addressed. Questions of precision and frequency are, of course, critical in computations of data collection costs both to the government and to respondents. The lower the precision and frequency requirements, the lower will be the costs because samples can be smaller and less frequently surveyed.

Seven data needs can be set forth. First, there is a clear need for a universe listing of nonpublic schools, to be used to draw samples for any study and to provide the basis for monitoring the rate of openings and closures. The technical problems in constructing this list are nontrivial both because the definition of "school" has not been resolved and because some schools would rather not be included on a list maintained by the federal government. The universe listing, in addition to name, address, and telephone, should include information needed for sample stratification. This information would include descriptors not likely to vary from year to year but that account for large differences in the activities and environment of the school, such as whether the school is elementary or secondary, what its religious affiliation is (if any), whether it serves boarding students, and whether it serves students of both sexes. It is not difficult to imagine studies for which these data would provide substantial increases in sampling efficiency.

Second, there is a need for enrollment data, so that increases or decreases can be included in economic models of the school system, models to be used in predicting the outcomes of federal policy alternatives.

Third, there is need for financial data, both for trends in per-pupil costs and expenditures and for levels of tuition and board paid by parents and students.

Fourth, there is need for information on student characteristics, including race, socioeconomic status, and abilities.

Fifth, there is need for information on the type and quality of services being provided to nonpublic school students: classroom size, hours of instruction, available facilities, course offerings, and

noninstructional services such as counseling. Information on accreditation of schools and the education and experience of teachers is also needed to determine quality of instruction.

Sixth, there is a need for information on participation in federal programs. This includes the amount of funding received, the numbers of students served, and the services provided to those students.

Seventh, there is a need for information on the relations of nonpublic schools to nearby public schools--the extent and types of communication, cooperation, and competition and the complementarity of the roles played by each.

Alternative Designs

These data needs suggest specification of a set of data collection efforts to be carried out over a period of, say, ten years, along with guidelines for adapting the plan to changing needs. The specification for each study must include (1) what types of information are to be collected--in terms of both the general area of data need to be addressed and the breadth and depth of measurement required, (2) what types of schools are to be included in the study, (3) how many schools are to be included in the study, and (4) the method of collecting the data. To determine which studies are needed, one lists for each of the seven data needs how precisely and at what times it must be filled. Then, among possible alternative solutions to filling the constraints, the design that minimizes cost can be selected.

There are eight questions that must be addressed in deciding on an optimal design. These will be considered in turn, and the results will be synthesized into a recommended design. Then specific measures will be compared, using the data from the three surveys, to determine which require frequent updates.

1. Is it really necessary to develop a nonpublic school data base?

The growth of interest in nonpublic schools nationwide and the increasing complexity of federal policies for private schools (mentioned in 86 different sections of the federal Education Amendments of 1978),

leave no doubt that policy development is occurring, and if resulting policies are to be responsive to the real needs of the schools, they must be based on sound data. The current data base, while substantial, is not nearly sufficient to perform the needed policy analyses, and the foundation must be laid for responding to the economic, political, and social changes that will occur in the 1980s. To ignore the need for this data base is unacceptable.

2. Is there sufficient reason for changing the design that has been used for the last three years?

The design of the current surveys was exploratory, and plans for the future should build on that exploration, replicating the strong points and improving the weak points. The stability of many items, such as school type and religious affiliation, indicates that these need not always be included. On the other hand, a universe survey requires too great a response burden to allow inclusion of the many indepth inquiries, such as in school finance and service quality, that are required for policy analyses. Studies based on samples selected representatively from the frame provided by the universe survey, will result in more efficient uses of taxpayer money and school staff time.

3. Is it really necessary to have a comprehensive, coordinated long-term plan?

A comprehensive, coordinated long-term plan is essential if the Education Department is to respond effectively to the congressional demand to eliminate redundant paperwork. Moreover, nonpublic schools will be more disposed to respond if they are assured that all government demands for data are limited to this single plan. By ensuring that each study relates to the same data base, the Education Department can convince researchers and policy analysts to refrain from duplicative efforts and provide for sharper analyses of the data that are collected. To develop support for the plan, researchers and policy analysts must take part in setting priorities and see that the comprehensive plan includes the opportunity to collect the data they need within the planning horizon.

4. Is it appropriate to maintain regular periodicity in data collection?

Although strict periodicity, such as maintained in the higher education general information survey, the triennial survey of participation in adult education, and the decennial national census, removes a computational step in the interpretation of trends, its importance does not compare with the importance of providing results when they are needed. Also, although periodicity does reinforce the arguments for funding "when the survey comes due," this is artificial and should be much less persuasive than conforming to a well-formulated and widely agreed upon long-term plan.

5. How often should the universe be updated?

The universe updates should be annual, partly for the purpose of building up government-nonpublic school cooperation. As described below, a postcard shuttle survey (in which respondents merely update computer generated responses from the prior year's survey), coupled with an annual newsletter describing federal policies and programs for nonpublic schools, could be made attractive to the potential respondents (and nonrespondents). The annual update can be made more attractive by including opportunities for respondents to state their views on current policy problems if they desire. Such a system could, in fact, become a model for cooperative survey programs. The value of periodic updates for the file development process is clear, because they are needed to ensure that the file will not become obsolete. The need for annual updates, as opposed to biennial or triennial, is to minimize the difficulty of the update. The cost of dealing with nonrespondents will exceed the mailing costs, and the number of nonrespondents will increase with the lag between updates.

6. Are longitudinal data collections necessary, or will cross-sectional studies on samples suffice?

The requirements for policy analysis frequently involve longitudinal data collection to assess the effects of policy interventions. Longitudinal data are significantly more expensive and less timely, however, than cross-sectional data, so collecting them should be avoided where

unnecessary. The three years of data on the nonpublic school universe provide a good basis for longitudinal analyses based on future cross-sectional data collection, if it is referenced to an integrated data base. Measures important for longitudinal analyses should be considered for inclusion on a universe update survey form as soon as feasible.

7. Are aggregate data, such as from Catholic dioceses and Seventh Day Adventist conferences, adequate?

Aggregate data are generally not sufficient for valid policy analyses. If respondents with one religious affiliation respond as individual schools while others respond as aggregates (e.g., dioceses, conferences), analyses can be made uniform only by imputing individual school responses randomly from the aggregated figures. In order to estimate effects of policies on schools accurately, it is necessary to have this uniformity. A compromise for the universe survey would be to gather information on every individual school once, afterward only on the aggregate unit, except for a small representative sample of individual schools. This would provide the basis for statistical imputation of the responses of other schools and validation of the imputation process. In any case, the data file must be maintained at the school unit of disaggregation to be useful for analyses.

8. Can the data base be developed without substantive contributions from federal policy planners and extensive design involvement of representatives of public and nonpublic schools?

The effort must be cooperative. The government's policy development and data collection functions cannot be separated without damage to both; and the nonpublic schools must be able to see measurable benefits for themselves, if they are to be convinced to respond. Nonpublic school administrators know more about the meaning of the data in the data base than government policymakers and data collectors, and their expertise should be used in planning.

Generally, in view of these eight issues, the best design appears to be a combination of (1) a single annual universe updating effort to provide the basis for drawing samples, (2) a few indepth case studies of a few dozen schools each to identify the most valid measures of each data type to be included in the survey, and (3) frequent studies of samples of a few hundred schools selected representatively from the universe to monitor the most changeable statistics and respond to particular information needs.

The universe update is necessary if one is to have any confidence that data from samples pertain to the whole of the nonpublic school universe. Two parts of that effort need be considered: (1) updating information on known schools and (2) identifying new schools. Using automation, the former can be done at a small cost, both to the government and to respondents, by sending out computer-generated forms displaying information to be returned with corrections. Sending this with a brief (two- to four-page) summary of information relevant to nonpublic schools (e.g., survey results and government policy activities) should ensure a high response rate. The latter effort, identification of new schools, requires creativity, however. The most inexpensive sources may be program lists, organizational membership lists, and state education agency lists. The feasibility and value of consulting commercial lists, such as those used by publishers, should also be considered. The key concept for efficiency is to identify some social system to which new schools will attach themselves, such as the telephone system, and through which they will announce their existence. Any such indirect method of identifying new schools should be validated, however, by direct study of a small sample of counties, to determine what proportion of new schools of various types are missed by the method.

The need for indepth studies in certain areas also is of prime importance. As pointed out in the Seminar on Meaning and Significance in the Analysis of the Nonpublic School Survey (McLaughlin, 1978), expenditure and income data are highly questionable and could be made substantially more meaningful by adopting terms for response that are less ambiguous and more universally used. Data on type and quality of services,

likewise, need refinement of terms and meanings. Information needs on the relationships between public and nonpublic schools need to be clarified, as no clear delineation of this information-type exists. At the present time, the need for indepth studies in areas other than these three is less clear.

Indepth studies can be conducted very infrequently, possibly only once or twice in 10 years for each topic, but it is essential that they be conducted early because the design of instruments for other studies will be enhanced on the basis of their results. Thus, we come to the conclusion that indepth studies in these three areas (finance, quality, and nonpublic/public school relations) should be carried out during the 1980-81 school year, if possible. The indepth studies should involve site visits with extensive data gathering to verify meanings behind objective responses. The sample size could be fairly small, merely sufficient to cover the typical variations among schools and to support calculations of correlations among measures, about four dozen schools for each area of concern. In order to minimize cost and at the same time to provide for studies of the interactions among finances, quality, and nonpublic/public school relations, the same schools and site visits could be used for all three areas. The primary problem with this solution is that it could put undue burden on a small number of schools. To compensate for this, the government could offer a substantial inducement for participation, such as support for some appropriate educational activity, much of the cost of which would be recovered through the increase in efficiency derived from the decrease in number of sites. In cost, the indepth studies would represent a high loading early in any long-range plan; however, any argument for delay, such as to spread out the funding commitment, must take into account the decreased value of the indepth studies after the frequent small studies are in progress. Refining measures after a few years rather than at the beginning of data collection will reduce the utility of the longitudinal data base.

A series of frequent small studies forms the third component of the design. These are studies that would use a short mailed questionnaire or a 10- to 15-minute telephone interview to gather measures to address sets of related

policy issues, such as in school finance. Using a probability sample based on the universe sampling frame, these surveys would aim for quantitative estimates of nonpublic school population characteristics. The size of each survey will be a function of the accuracy required by policy developers; and to ensure that estimates of required accuracy have a rational basis, discussions with the policy developers should identify the minimum differences that would be policy-relevant.

For example, suppose a study of public-nonpublic school interactions aimed, among other things, to find correlates of increases of children in nonpublic schools. It might not be very important to be able to identify relations that accounted for less than 3% of the total variance, but relations that account for more than 4% of the variance might be of interest. This requirement would suggest a sample of about 100 schools. If the same study were used to identify several relations or to identify differences between subgroups, a larger sample would be necessary. In general, samples from 200 to 800 are likely to cover most needs for precision of estimates. Even if four such surveys are performed each year, each individual school in the universe would, on the average, participate in only one during the decade.

The cost of the surveys will depend greatly on the startup costs of each. It would be most efficient to carry them out through the same survey mechanism, within a nonpublic school information program at the Center. This would allow for repeated use of the same sampling algorithms, the same instrument development procedures, the same analysis programs, and the same liaison with nonpublic school representatives. Moreover, by coordinating the sampling across surveys, as by drawing samples for several studies simultaneously, the government can balance the response burden evenly over the universe of nonpublic schools.

An important but inexpensive supplement to this data collection plan is short-term year-to-year updating of information on the entire universe using statistical prediction equations. The resulting "imputed" file would be expected to diverge from actual values as random error accumulates, especially for volatile measures. By adjusting

the imputation formulas on the basis of the sample survey outcomes and the universe update, however, the divergence can be reduced.

Empirical Results

The data from the three years of NCES surveys can be used to estimate the constancy and predictability of the variables measured, and this constancy, in turn, is an indicator of lack of need for frequent reassessment. For each measure, we can address the following questions:

(1) How variable are the responses from one year to the next?

Change in ratio scale variables, such as the measures in this survey, can be measured as percentages of the values for one year. Although not presented here, one might estimate significance of changes as indicators of reliable trends; for example, the year-to-year change might be compared to within-year variation.

(2) How predictable are the responses from the prior year's responses?

Variables that change from year to year may still not need frequent data collection if the values are highly predictable from the prior year's data. A secondary consideration is necessary, however:

(2') How variable is the prediction function from one year to the next?

With three years' data, the information is minimally sufficient to estimate the variability of the prediction function: how different is the 1978 prediction function based on 1977 data from the 1977 prediction function based on 1976 data? If the prediction functions change, then the value of any prediction function is lessened.

It is very hazardous to extrapolate from three data points (1976, 1977, and 1978, for example) to estimate values in future years. Many things can happen to produce discontinuities in the time series. It is possible, however, to set reasonable lower bounds on our uncertainty, based on assumptions that changes in the future will be about as

predictable as the changes over the observed period. Using these assumptions, we can estimate when a change of, say, greater than 10% from the current value is likely to occur. An argument could be made for reassessing a variable when it is projected to have changed more than a particular threshold, with that threshold dependent on the importance of the variable for policy research. Early detection of a 1% change may be called for in a very important variable, such as total enrollment, because, for example, of its implications for the demand for new teachers. (A 1% increase in enrollment may necessitate a 30% increase in new teacher certifications, if certifications are otherwise at a rate just sufficient to replace retiring teachers.) Early detection of a 25% change may, on the other hand, be sufficient for a less important measure, such as the number of elementary boarding schools with a particular religious affiliation.

If one sets a 10% change cutoff, then those that change less than 2% per year may not need to be assessed again until five or more years have passed, while those that change by more than 10% annually should be assessed every year. In order to provide a summary of the data relevant to this issue, we have presented the measures, classified by rate of change, in Table A-1. The percentage change was estimated as the average for the two available year-to-year transitions. These are minimal data for such analyses and alternative interpretations exist of the year-to-year changes; however, these summaries provide the information relevant to the question of frequency reassessment that can be extracted from the three-year survey at present.

The results presented in Table A-1 indicate that most of the measures are relatively slow-changing. Only two of the charted measures changed on the average more than 5% per year, and the fact that these were for the smallest group, Lutheran schools, suggests that they may not be significant. Per-pupil expenditures, corrected for inflation, tended to change more than numbers of schools, students, teachers, graduates, or federal program participation. Expenditure data were, however, the least reliable of the measures included in the survey, and any conclusion based on these data should be viewed as extremely tentative.

Table A-1

Classification of Selected Measures as Fast-Changing or Slow-Changing[a]

	≤ 2% per year	2% to 5% per year	>5% per year
Number of schools			
Catholic	✓		
Lutheran	✓		
Other affiliations	✓		
Non-affiliated		✓(-)[b]	
Elementary & Middle	✓		
Secondary	✓		
Combined	✓		
Other		✓(-)	
Number of teachers			
Catholic	✓		
Lutheran	✓		
Other affilications	✓		
Non-affilated	✓		
Elementary & Middle	✓		
Secondary	✓		
Combined	✓		
Other	✓		
Number of students			
Catholic		✓(-)	
Lutheran	✓		
Other affiliations		✓(+)	
Non-affiliated	✓		
Elementary & Middle	✓		
Secondary	✓		
Combined	✓		
Other			✓(-)

(Continued)

Table A-1 (continued)

	≤ 2% per year	2% to 5% per year	>5% per year
Per-pupil current operating expenditure[c]			
Catholic	✓		
Lutheran		✓(-)	
Other affiliations		✓(-)	
Non-affiliated	✓		
Elementary & Middle		✓(+)	
Secondary	✓		
Combined		✓(-)	
Other		✓(-)	
Number of graduates			
Catholic	✓		
Lutheran			✓(-)
Other affiliations	✓		
Non-affiliated	✓		
Title I participating schools			
Catholic		✓(-)	
Lutheran			✓(+)
Other affiliations	✓		
Non-affiliated	✓		
Title IVB participating schools			
Catholic	✓		
Lutheran		✓(-)	
Other affiliations	✓		
Non-affiliated	✓		
Elementary & Middle	✓		
Secondary	✓		
Combined	✓		
Other	✓		

[a]This and other displays in Appendix A were prepared before reimputation of missing data on the nonpublic school survey file and may differ somewhat from results presented elsewhere in the report.

[b]Positive signs indicate an increase; negative signs, a decrease.

[c]Changes are classified as deviations from the overall inflation rate of 7% per year.

The predictability of selected measures from the prior year is shown in Table A-2, taken from the first version of the imputation reported in May 1979 (McLaughlin & Wise, 1979). This table is supplemented by information present in Appendix D on the accuracy of the imputation method. Generally, the data are highly predictable from the preceding year.

Finally, there is the question of how variable the prediction function itself is, from one year to the next. We have addressed this for three variables: total enrollment, total classroom teachers, and total expenditures, by comparing the predictions of 1978-79 responses from 1977-78 responses using both the 1977-78 to 1978-79 function and the 1976-77 to 1977-78 function. In particular, the two predictions were compared in terms of their estimates of differences between respondents and nonrespondents and in terms of the correlations between estimates for individual schools.

The mean 1978 school enrollment estimates for nonrespondents imputed from earlier data were 157 using the 1977 to 1978 function and 167 using the 1976 to 1977 function, compared with the average of 253 for all 1978 cases; and the correlation between the estimates from the two functions was .86. For mean number of classroom teachers, the two estimates for nonrespondents were 10.6 and 11.7, compared to the overall average of 13.8; with a correlation of .79. For mean school expenditures the corresponding estimates were $188,000 and $191,000, compared to the overall average of $244,000; with a correlation of .90. For these three variables, then, the imputation functions using the prior year's data varied little between 1976 to 1977 and 1977 to 1978. Thus, tentative population estimates for 1979-80 might well be obtained as projections from the 1978-79 responses, assuming school closures and openings were updated.

To summarize, the empirical results indicate that changes are occuring from year-to-year and that they are not completely predictable. On the other hand, they are not chaotic changes, and collection of data on samples every year or two should be sufficient to develop valid population estimates and projections.

TABLE A-2

Evidence for Potential Imputations of 1977-78 Private School Data from 1976-77 Responses, for Selected Items

	(A) Number Cases Missing 1977-78 Data	Percent of Those in (A) that Have 1976-77 Data	Percent of Those with Data at Both Points that Have the Same Responses at Both Points
Day-Boarding Mix (3 Alternatives)	826	69%	98%
Sex-Served (3 Alternatives)	818	70%	98%
School Type (7 Alternatives)	802	70%	95%
Population Center Served (6 Alternatives)	800	70%	75%[a]
Foreign Language Courses (2 Alternatives)	5054	29%	97%
Title I Assistance (2 Alternatives)	5066	29%	90%
Title VB Assistance (2 Alternatives)	5059	29%	90%
Title IVC Assistance (2 Alternatives	5078	30%	87%
School Lunch Assistance (2 Alternatives)	5058	29%	93%
Other Federal Assistance (2 Alternatives)	5089	30%	89%

[a] 63% of the remainder are within one rank on this ordinal scale.

References

Russ-Eft, D. F., Rubin, D. P., & Holmen, R. E. Identification of Issues in adult education (SAGE Technical Report No. 1). Palo Alto, Calif.: American Institutes for Research, 1979.

APPENDIX B

Nonpublic School Survey Form

Notes:

1. This is the form for 1978-79, for non-Catholic schools. The form for Catholic schools omits item 11 and item 14 breaks down classroom teachers into the categories of religious and lay.

2. The 1977-78 forms are essentially identical to the 1978-79 forms. The 1976-77 forms included sex breakdowns of the grade-by-grade enrollment figures and did not separate tuition and fee income from other sources.

3. Circles indicate items administered in telephone follow-up of nonrespondents.

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
NATIONAL CENTER FOR EDUCATION STATISTICS
Washington, D.C. 20202
SURVEY OF PRIVATE ELEMENTARY AND SECONDARY SCHOOLS

Form Approved
OMB No. 51-R1199

This report is authorized by law (20 U.S.C 1221e-1). While you are not required to respond, your cooperation is needed to make the results of this survey comprehensive, accurate, and timely. When you have completed this form, please return it according to the instructions contained in the letter which accompanied the form. The data reported on this form will be disclosed upon request to the public under the requirements of the Freedom of Information Act (5 U.S.C. 552).

1. NAME AND ADDRESS INFORMATION

A. Name of School

B. Telephone (include area code) ( )

C. Street address, P.O. Box, or RFD number

D. City

E. State

F. ZIP

2. LOCATION OF SCHOOL

A. Name of County

B. State

3. Number of the congressional district in which this school is located

4. How would you best describe the population center where this school is located?
A. [ ] 500,000 or more population
B. [ ] 250,000 but less than 500,000
C. [ ] 100,000 but less than 250,000
D. [ ] 50,000 but less than 100,000
E. [ ] 2,500 but less than 50,000
F. [ ] Less than 2,500 or rural area

5. Is the student body composed o
A. [ ] boys only
B. [ ] girls only
C. [ ] boys and girls

6. Are the students
A. [ ] day only
B. [ ] boarding only
C. [ ] mixed day and boarding

7. How would you best classify this school by type? (check only one)
A. [ ] ELEMENTARY - any combination of grades lower than 9. Does not include any school classified as MIDDLE or junior high schools which should be classified as SECONDARY.
B. [ ] MIDDLE - any combination of upper elementary and/or secondary grades, generally organized to include at least three grades beginning with grade 5.
C. [ ] SECONDARY - any combination of grades above grade 6, including junior highs
D. [ ] COMBINED ELEMENTARY-SECONDARY - any combination of elementary AND secondary grades (unless classified as a middle school).
E. [ ] SPECIAL EDUCATION - serves handicapped students exclusively. Such schools must offer curricula and services designed to meet the needs of the physically handicapped, emotionally disturbed, learning disabled, and/or mentally retarded.
F. [ ] VOCATIONAL TECHNICAL - serves vocational/technical students exclusively. Such schools must offer curricula and programs in one or more semi-skilled, skilled, or technical occupations designed to permit persons who complete to seek employment in the field of training.
G. [ ] ALTERNATIVE - not an adjunct to or part of a regular school. Provides nontraditional educational programs designed to meet needs of students which generally cannot be met in the regular school.

NCES FORM 2325 (Rev. 7/78)

8. Check each of the programs listed below which are offered in this school.

[ ] a. Regular academic
[ ] b. Special education for the handicapped
[ ] c. Vocational technical education
[ ] d. Compensatory education for the disadvantaged

9. Does this school offer any courses other than foreign languages which are taught in a language other than English?

[ ] a. YES [ ] b. NO
(if yes, specify the language)

10. Do students in this school participate in activities or programs or use materials, supplies, or equipment supported by any of the federal assistance programs listed below? (check yes or no for each program listed)

	YES	NO
A. Elementary and Secondary Education Act of 1965 (ESEA)		
a. Title I, educationally disadvantaged children		
b. Title IV, Part B, school library resources, equipment, testing, guidance, etc.		
c. Title IV, Part C, supplementary centers and services, exemplary programs, dropout prevention, etc.		
B. Federal school lunch and/or special milk programs.		
C. Other federal assistance programs (specify below)		

11. Is this school affiliated with a religious group or organization?

[ ] a. NO [ ] b. YES (if yes, check the appropriate affiliation below)

[ ] a. Baptist
[ ] b. Calvinist
[ ] c. Eastern Orthodox
[ ] d. Episcopal
[ ] e. Friends
[ ] f. Jewish
[ ] g. Lutheran
[ ] h. Methodist
[ ] i. Presbyterian
[ ] j. Seventh Day Adventist
[ ] k. Other (specify below)

12. If this is a secondary school, how many students graduated from the 12th grade during the 1977-78 school year, including the summer of 1977?

Boys	Girls	Total

13. How many students were on the official roll of this school on October 1, 1978 or the nearest date thereto when enrollment was considered to have stabilized?

Grade	Number	Grade	Number	Grade	Number
PK		4		9	
K		5		10	
1		6		11	
2		7		12	
3		8		Unclassified	
				Total	

14. What is the full-time equivalent number of persons employed to work in this school on or about October 1, 1978? Do not include vacant positions and volunteer personnel who work without pay.

Full-time equivalence is defined as the amount of time a person spends in an assignment category divided by the amount of time a full-time employee would spend in the assignment category. One full-time equivalent is expressed as 1.0. The total full-time equivalence for each category should be rounded to the nearest whole number. The following examples illustrate how full-time equivalence should be computed.

a. If a teaching position were filled by two persons, one male and the other female, each working one-half day, count each person as 0.5, adding 0.5 to the count for males and 0.5 to the count for females.

b. If a full-time employee were assigned to teaching one-half time and to counseling one-half time, count this person as 0.5 teaching and 0.5 other professional.

Assignment Category	Male	Female	Total
A. Classroom teachers			
B. Other professionals, e.g., librarians, principals and deans, counselors, etc.			
C. Other personnel, e.g., food service workers, clerks and secretaries, etc.			

15. Financial Information for the 1977-78 School Year

a. Be sure to include all subsidies, gifts, and earnings on investments under "Other Income".
b. Current operations include income and expenditures for the operation of the academic program of the school. Expenditures should include those salaries and other expenses for personnel, administration, operation and maintenance, instructional materials and supplies, etc.
c. Auxiliary operations include income and expenditures for operation of dormitories, bookstores, summer camps, transportation, food services, etc.
d. Do NOT include expenses for acquisition of capital assets or debt service.

Purpose	Income			Total Expenses
	Tuition/Fees	Other	Total	
	(1)	(2)	(3)	(4)
A. For current operations				
B. For auxiliary operations				
C. Totals (sum of lines A and B)				

NOTE: Please use the remainder of this form to comment on any item(s) you found lacked clarity or were difficult to understand. If you have any suggestions which might improve this survey in future years, please make them.

APPENDIX C

File Development and Merging

Overview

In each of the three years, 1976-77, 1977-78, and 1978-79 (referred to respectively as 1976, 1977, and 1978 herein), the survey questionnaires were collected by the National Catholic Education Association (NCEA) and the Council for American Private Education (CAPE), where preliminary editing was performed. Forms were forwarded to NCES and submitted to the DHEW Data Management Center for transfer to tape. At NCES the Institutional Surveys Branch edited the resulting computer files to reduce errors in coding. The files were then turned over to SAGE for imputation of missing data. SAGE determined that merger of the data from the three years was an essential first step in the imputation process.

Although the content and record layout of the edited files were nearly identical for the three years, there were several significant differences among the files.

1. Except for a few dozen errors that SAGE corrected, the 1977 and 1978 files had the same school code number descriptions, facilitating a longitudinal merge. The merge with 1976 data, however, required matching of free-field names and addresses as well as zip codes and telephone numbers.

2. No records of nonrespondents were included on the 1976 file, which presented a problem for identification of 1977 school openings. Nonrespondents were included on the 1977 and 1978 files.

3. Seven hundred individual Seventh Day Adventist (SDA) schools were included on the 1976 file, 67 of which were marked as not affiliated with any religious group. No SDA schools were included in the 1977 and 1978 files, although aggregate responses from 52 SDA Conferences were received, representing 1,047 schools in 1977 and 1,176 schools in 1978.

4. About 400 Catholic schools were not included on the 1977 file, although there were records that they had responded in all three years.

5. Several hundred duplicate records were present on the 1976 file, and a smaller number of duplicates were present on the 1977 file.

Several other imperfections in the files were discovered and corrected.

NCES realized the need for systematic development of the file and delegated that task to SAGE. During the summer of 1979, staff of SAGE and the Institutional Surveys Branch worked together to enhance the file. Specifically, the task assigned to SAGE was to impute missing data, first for the 1977 file, and then, in order to provide the basis for three-year trend analyses, for the 1976 and 1978 files. A first version of imputation of 1977 data was summarized in a draft report (McLaughlin & Wise, May 1979). Completion of the imputation of missing data required substantial additional editing of the file by SAGE, and some needs for editing remain.

SAGE divided the imputation task into two phases: (A) generation of a merged three-year data file with a single record, containing all available data, for each school that responded at least once and was not later ruled out as not a member of the universe (several hundred schools were deleted as being preschool or as being public, and a few schools were deleted when it was decided to exclude the outlying territories from the file); and (B) imputation of missing data on each record. Correct assignment or imputation of open-closed status codes for each school is, of course, crucial for the validity of analyses based on the file, and it received special attention in the processing, as described below. The order of steps is summarized in Figure C-1. The merge processes are described below, and the imputation processes are described in Appendix D.

Merging Single-Year Files

The first step was merger of the 1976 and 1977 files. Using computer matching on various combinations of name, address, telephone number, city, zip code, and school type, some 14,500 of the 18,100 records on the 1977 file were matched to records on the 1976 file, leaving about 2,600 1976 records unmatched. These 2,600 cases were examined in detail and were assigned to categories as shown in Table C-1. There were about 800 cases for which the open-closed status in 1977 was not clear. For

these 800 cases not identified as either SDA schools, matches to previously unmatched 1977 cases, duplicates, or Catholic schools whos responses were lost in processing, it was necessary to assign or impute an open-closed status code for 1977; and three steps were taken to accomplish this. First, directory assistance was used for those with telephone numbers on the 1976 file (about 90%). Of 742 cases, 472 had telephone numbers in 1979 and so were tentatively assumed open in 1977-78, and the other 270 were assumed closed. Second, schools marked as closed in NCES's 1977 receipt log were coded appropriately on the file. Third, a sample of 100 of the 800 schools was contacted to validate the directory assistance assignments. Thus, all cases with responses in either 1976 and 1977 were assigned a 1977 open-closed status code.

The next step was to determine how many of the remaining 2,500 cases with 1977 data but no 1976 data were open in 1976. In order to estimate a base rate for distinguishing between schools that were new in 1977 and schools that existed but did not respond in 1976, a sample of 100 was called, and seven were found to have been new schools in 1977. Other 1977 measures were then used in arriving at an imputed 1976 open-closed status for those not called. This is described later.*

As soon as the 1978 file became available, in late August 1979, the 1978 data were merged with the longitudinal 1976-77 data. The process by which the final count of 20,504 cases was reached is summarized in Table C-2.

---

*The open-closed status codes for 1976 were, in fact, the final variable imputed, in order to make use of maximal information for this crucial variable. Thus, imputed values exist on the file for schools in the year prior to their inceptions for use if these status codes are changed. The file to be released for use will have data on closed schools set to missing to avoid erroneous analyses.

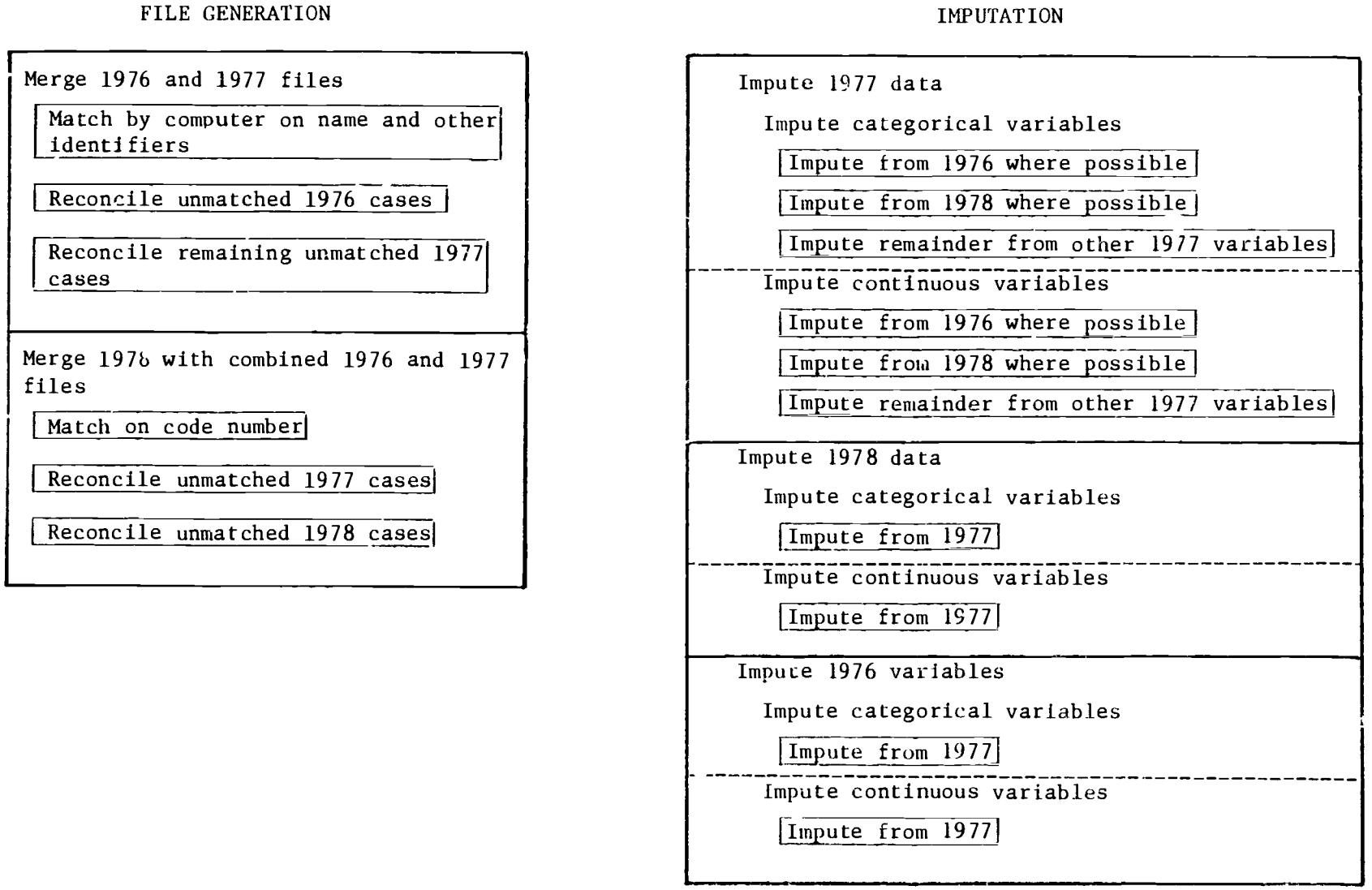

Figure C-1. Overview of the file development performed by SAGE.

Table C-1

Ultimate Disposition of Cases Appearing on
1976-77 File but Not Matching 1977-78 Cases

	No.	Original Unmatched 1976 Cases
a.	657	SDA Schools
b.	67	Additional SDA schools (coded as non-affiliated on the 1976-77 file)
c.	191	Duplicates of other unmatched 1976 cases
d.	622	Hand matches (440 of which were duplicates)
e.	400	"Lost" Catholic school records, imputed as "open"
f.	111	Listed in survey log as "closed"
g.	213	Subsequently removed as duplicate or "inappropriate" schools
h.	386	Open/closed status in 1977-78 imputed randomly according to base rate
	2,647	

Table C-2

Source of Cases Included in the Final File of Nonpublic Schools

Number	Changes
18,103	1977-78 cases (after removing initial duplicates) as of July 1979
1,047	Additional SDA schools in 1976-77 and 1977-78
656	Additional nonrespondents in 1977-78
1,110	Unmatched 1976 cases (e, f, g, and h in Table C-1)
-47	Duplicates noted in August 1979
-14	Outlying Territories
-473	Deletions of duplicates and inappropriate schools
-91	Additional duplicates found in final two merges
-92	Additional duplicates found in September 1979
155	New cases from 1978-79 Master
150	New 1978-79 SDA schools
20,504	

APPENDIX D:

Description and Validation of the Missing Data Imputation Procedures

The missing data imputation procedures used for the 1976-1978 Nonpublic School Surveys were organized into four "passes." The first two passes imputed missing values for discrete variables (school characteristics), while the third and fourth passes handled many-valued variables (enrollment, graduates, staff, and budgets). Passes one and three processed cases where data were available from the prior year's survey; passes two and four used procedures for imputing missing values from other variables in the same year's survey for cases where prior year data were not available. The validation of the imputation procedures used for the 1976-1978 Nonpublic School Surveys was organized into separate substudies of the procedures used in these four passes.

The missing data imputation procedures had been applied first to the 1977 data (1977/78 survey) using other years' data* where available. After all 1977 data were imputed, the procedures were applied to the 1978 data (1978/79 survey) using the actual or imputed 1977 values as the prior year's values. Then, the missing 1976 values were imputed using the 1977 values as "prior year" variables. In each step after the first, prior year's values were always available, since all 1977 values were filled in for each case in the first step. As a consequence the second and fourth passes described above were only used in the imputation of 1977 data.

The validation study focused on the imputation of the 1977 data, because this provides a sufficient basis for evaluating the procedures used for all three years. The method of validation was to use the imputation procedures on cases not requiring imputation. For these cases, distributions of imputed values could then be compared to distributions of

*A few schools with missing 1977 values had reported values for both 1976 and 1978. In this instance, the 1976 values were used as the basis for imputation.

reported values. Following the suggestions of advisors, we performed the validation comparisons (except for Pass 1, below) separately for subpopulations different from and similar to real nonrespondents. A finding that the procedures work as well for populations similar to real nonrespondents as for others is important, because we have no way, short of further data collection efforts, of fully validating the imputations on the real nonrespondents themselves.

Pass 1: Imputing Discrete Variables (Questions 4 through 11) from Prior Year's Values

The procedure used in Pass 1 was simply to replace a missing *discrete* variable with the prior year's value. The validation of this procedure consisted of an examination of the extent to which schools reported the same value for each discrete variable on consecutive surveys. Table D-1 shows the percentage of schools with data in both the 1976 and 1977 surveys that reported the same value in both years. The high degree of consistency in these figures indicates the validity of this approach. The only variable showing any real degree of inconsistency is the question regarding the size of the population center in which the school was located. Since very few schools actually moved, this inconsistency would seem to indicate a lack of reliability in the reported responses rather than any real error in the procedures. It should also be noted that fewer than 10% gave responses that differed by more than one category on this ordinal scale. Even so, a separate procedure was used for this variable because of the lower level of agreement between the two survey years. (See the description under Pass 2.)

In general, a simple replacement procedure will underestimate the degree of change in the variables being replaced (other than change due to new and closed schools). In the present case, however, the degree of change in established schools is so small as to be negligible, since apparent shifts out of a category were generally matched by a roughly equal number of apparent shifts into the category. Table D-1 also shows the largest "net" shift in any category. A small apparent decrease (1%) occurred in the number of coeducational schools, and small increases occurred in the number of schools indicating support from specific federal

TABLE D-1

Percentage of Cases Providing Responses to Successive Surveys Giving the Same Response (for Each Discrete Variable)

Variable	76/77-77/78	Maximum Change in Percent in a Category
Sex Served	98.4	1.0
Day or Boarding	97.8	.5
Programs		
Academic	99.2	.0
Special Educatic	97.7	.0
Vocational/Technical	97.7	.0
Compensatory	95.8	.0
Some classes taught in a language other than English*	97.5	.7
Federal Support		
Title I	90.0	1.9
Title IV-B	90.1	1.7
Title IV-C	86.6	2.2
Lunch/Milk	93.4	2.7
Other	88.8	-1.1
Population Size of Community**	74.6	1.8

* The question was worded, "Does this school offer any courses other than foreign languages which are taught in a language other than English?"

** Percent differing by 1 category = : 16.0

programs. In the latter case, the decrease in the percentage of schools indicating "other" federal support suggests that the apparent changes may be due partly to reporting inconsistencies.

Pass 2: Imputing Discrete Variables from Other Variables

The procedure used to impute discrete variables from other variables in the same survey was to compute the distribution of the variable to be imputed conditional on the value(s) of discrete variables present or already imputed and then randomly to select a value for the missing variable with probabilities matching the appropriate conditional distribution. The variables were ordered such that those that had the least amount of missing data (those included in the telephone survey) were imputed first and then the values of these variables (either real or imputed) were used in the imputation of the variables with more frequent missing data. First religious affiliation was set randomly for the three cases (out of 20,504) in which it was missing, 50% to "no affiliation" and 50% to "other religious affiliation." The second variable, school type, was imputed randomly according to base rates within religious affiliation (recoded into four categories). For the next two variables, sex of students and day or boarding, conditional distributions were computed for each combination of religious affiliation and school type (also recoded to four categories, yielding 16 separate conditional distributions).*

The remaining questions (8a through 10e) were all dichotomous. Each of these variables was regressed on a set of dichotomous indicators based on the preceding variables, as shown in Table D-2. All of the cases in the file were then split into eight strata based on the value of the regression function. (To ensure that observations would not be clustered in a single stratum, a cell width of .5 times the standard deviation of the regression function value was used to create eight cells with roughly commensurate expected frequencies.) A conditional distribution of the

*One other variable, the population center size, was imputed separately using a "hot deck" procedure that set each missing value to the value of a case with the same (or nearest) zip code. As mentioned above, the reliability of the reported values was in some question, so this variable was not included in the validation study.

TABLE D-2

Predictor Variables for Dichotomous Measures

Predictor	Characteristics
1	School is either secondary or combined elememtary-secondary
2	School is either elementary or combined elementary-secondary
3	School is one of the special other types
4	School is large (enrollment > 300, teachers > 10 or graduates > 50) (predictor = .5 if all missing)
5	School is coeducational
6	School is for girls only (predictor = .5 if coeducational)
7	School is for day students only
8	School is Catholic
9	School has no religious affiliation
10	School is Lutheran
Target*	
11	School offers regular academic program
12	School offers special education program
13	School offers vocational education program
14	School offers compensatory education program
15	School offers nonlanguage course in language other than English
16	School receives Title I assistance
17	School receives Title IVB assistance
18	School receives Title IVC assistance
19	School receives school lunch assistance/school receives other federal assistance
20	School receives other federal assistance

*Note: Each target variable, once imputed, was used as an additional predictor for later target variables.

(1) the overall distributions of each variable,

(2) the extent to which relationships among the discrete variables were reproduced, and

(3) the extent to which the actual and imputed values matched on a case-by-case basis.

As a further validation, the 1977 nonrespondent sample with available 1976 data was used. For these schools, the values of the 1977 variables that had been imputed from Pass 1 (i.e., from the 1976 data) were masked and then reimputed using Pass 2 procedures. The comparisons that were made were the same as those made for the mail and telephone respondents, except that the 1976 values were used as the "actual" values. In this case, however, mismatches could be attributed to errors in either imputation procedure. This method of validating the Pass 2 imputations was deemed appropriate, however, because of the high degree of correspondence between the 1976 and 1977 values shown in Table D-1.

Table D-3 shows the actual distributions for the mail and telephone samples for each discrete variable and the distributions imputed using Pass 2 procedures for these same cases. The degree of match is surprisingly high, considering that the imputations were almost totally blind. The imputation of school type, for example, was based only on the religious affiliation of the school. The mail and telephone samples differed widely in their school type distributions, the mail sample containing 75% elementary schools while the telephone sample contained only 40% elementary schools. Using only information of the different religious affiliation distributions for the two samples, the imputation procedures yielded estimates of 73% and 47% for these two samples. Similarly, using only information on the religious affiliation and (imputed) school type for each school in the two samples, the imputation procedures matched the difference in distribution of the sex served variable for the two samples. (Note that the mail survey contained slightly more "girls only" schools than "boys only" schools, while the telephone sample contained more than twice as many "boys only" schools as "girls only" schools.)

TABLE D-3

Actual and Imputed Distributions for Each Discrete Variable by Sample

Variable	Mail Survey		Telephone Survey	
	Actual (n=1066)*	Imputed	Actual (n=988)*	Imputed
School Type				
A. Elementary	74.6	73.0	39.7	47.3
B. Middle	.6	.7	.5	.7
C. Secondary	13.9	13.0	8.8	9.7
D. Combined	7.2	9.4	36.0	31.4
E. Spec. Educ.	2.8	3.0	12.5	8.1
F. Voc/Tech	.1	.1	.0	.0
G. Alternative	.8	.9	2.5	2.8
Sex Served			(n=980)	
A. Boys Only	4.0	3.6	4.2	4.8
B. Girls Only	4.2	4.2	2.0	2.0
C. Coeducational	91.7	92.2	93.8	93.2
Day or Boarding			(n=976)	
A. Day Only	95.4	94.8	88.4	86.9
B. Boarding Only	.8	1.5	4.7	3.8
C. Day and Boarding	3.8	3.8	6.9	9.3
Programs Offered				
A. Regular Academic	97.8	97.2	Insufficient Data	
B. Special Educ.	4.5	4.3		
C. Voc/Tech	2.4	2.2		
D. Compensatory	3.9	4.1		
Some classes taught in a language other than English	2.7	3.3		
Federal Programs				
Title I	40.2	39.6	Insufficient Data	
Title IV, B	82.7	81.4		
Title IV, C	13.9	13.4		
Lunch/Milk	66.2	72.8		
Other	10.2	8.6		

*These are systematic, unbiased samples of one in twelve of the 12,849 mail and one in four of the 3,962 telephone respondents in 1977.

Table D-4 shows the same comparison of "actual" and imputed distributions for the nonrespondent sample. The degree of match for this sample was reasonably good, although not as good as for the respondent samples. The degree of match for the "federal support" variables, in particular, was somewhat lower than for the others. In general, the imputation procedures underestimated the differences between the nonrespondents and the respondent samples. This was because the conditional distributions that were used in imputing these variables (conditional on a regression equation based on the preceding variables) were estimated solely from the mail survey respondents. Therefore, the degree to which the imputations matched the differences between the mail respondents and the nonrespondents was actually quite comforting. Moreover, it should be remembered that the use of 1976 data as the criterion for the nonrespondent sample did introduce some noise. The data in Table D-1 suggest that the actual 1977 values for this sample may have been slightly higher than the 1976 values shown as "actual" in Table D-4.

The next step in the validation of the Pass 2 imputations was to compare actual relationships among the discrete variables with the relationships estimated from the imputed data. These relationships are important for research studies using the data file. Relationships among the discrete variables were summarized in terms of key row percents (conditional proportions) from selected two-way tables. Table D-5 shows the actual and imputed values for the row percents from six such two-way tables: day or boarding, special education programs, and Title I funding, each with school type and sex served. These relationships are reproduced in the imputed data to a high degree, with one glaring exception--the actual differences in the percentage of "day only" schools between schools with different values on "sex served" was not reproduced at all. The reason for this one failure stems from an exception in the procedures used in imputing these particular variables. In all other cases, the discrete variables were imputed from conditional distributions based on all of the preceding variables (including imputed values where necessary) Only by including prior imputed values in the prediction of missing values could we hope to reproduce the relationships of the imputed variable to the previously imputed variables. In most cases, the variable to be imputed

TABLE D-4

Prior Year's and Imputed Distributions for Each Discrete Variable Among Nonrespondents with Prior Year's Data

(n=2805)	"Actual"*	Imputed
School Type		
Elementary/Middle	59.6	54.1
High School	11.3	11.2
Combined	22.8	28.6
Other	6.4	6.1
Sex Served		
Boys	3.3	3.7
Girls	2.5	3.1
Coed	94.2	93.1
Day or Boarding		
Day Only	92.4	90.1
Boarding	3.0	2.3
Mixed	4.6	7.7
Programs Offered		
Regular Academic	96.1	96.6
Special Educ.	7.4	6.2
Voc./Tech.	4.5	2.0
Compensatory	4.4	3.2
Some classes taught in a language other than English	5.3	10.6
Federal Support		
Title I	17.9	26.9
Title IV, B	40.5	66.3
Title IV, C	5.7	10.5
Lunch/Milk	37.5	55.4
Other	5.8	6.7

*Actual data are for 1976-77. The sample consisted of the 2,805 schools for which no 1977-78 data were available but for which 1976-77 data were available. The 1976-77 values were used to fill in missing 1977-78 values for these cases, but as a validation of Pass 2, these 1977-78 values were also imputed using only other 1977-78 data.

TABLE D-5

Actual and Imputed Relationships among the Discrete Variables*

Row Category		Percent of Schools in Each Row That Indicate Day Only**			Percent of Schools in Each Row That Indicate Spec. Educ. Programs***			Percent of Schools in Each Row That Indicate Title I Funding***		
		Actual n	Actual %	Imputed %	Actual n	Actual %	Imputed %	Actual n	Actual %	Imputed %
A. School Type										
A,B	Elem./Middle	1190	98.5	97.8	801	1.6	1.3	801	49.1	47.8
C	Secondary	234	76.1	76.1	149	2.0	2.7	147	8.8	16.7
D	Combined	427	94.6	91.2	77	2.6	2.6	77	16.9	7.0
E-G	Other	188	65.4	57.3	41	73.2	73.2	40	22.5	40.5
B. Sex Served										
A	Males Only	84	54.8	89.4	44	9.1	13.6	43	14.0	21.1
B	Females Only	65	83.1	89.2	46	0.0	2.2	46	2.2	13.0
C	Coed	1892	94.1	91.2	978	4.5	4.0	976	43.1	41.6

*The relationships are summarized in terms of row percents. In the first cell, for example, 98.5% of the elementary/middle schools were "day only" schools while only 76.1% of the secondary schools were "day only."

**Based on the combined mail and telephone survey samples.

***Based on the mail survey sample only.

was regressed on the preceding variables. For school type, sex served, and day or boarding, however, a slightly different approach was used because little information was available to use to predict these first variables. First, conditional distributions for school type were estimated for each religious affiliation grouping (Catholic, Lutheran, Other, None) since this was the only predictor available. Then, conditional distributions for sex served and day or boarding were estimated independently of each other for each different combination of the religious affiliation and school type grouping variables. Because neither of these two variables was used in the prediction of the other, the relationship between them was reproduced poorly in the imputed data. Had the consequences of independent imputation of these two variables been known, they would not have been imputed independently.

It is important to distinguish between the degree of error in the validation sample and the corresponding degree of error in the total file. The validation sample was drawn to evaluate the procedures used for imputations and some aspects of the imputation procedures were found to introduce statistically significant biases. In many cases, however, these procedures were only applied to a very small proportion of the file so that the biases introduced are not of any practical significance for the file as a whole. For example, because the sex served and day-boarding questions not only were included in the telephone survey but also were available for a high proportion of the 1976 and 1978 cases, these variables were imputed using Pass 2 imputations for fewer than 4% of the cases (772 of the 20,075 schools open in 1977). Consequently, even though there was an oversight in the imputation procedure, the current data file is not seriously in error even for this one relationship.

Finally, Table D-6 shows the rate of agreement between the actual and imputed values on a case-by-case basis. While these rates do not seem particularly impressive when compared with predicted agreement based on marginal distributions, it must be remembered that the only information on which such correspondences could be based was the school's religious affiliation. To simulate total nonrespondents in the validation study, all of the reported values for each school, other than religious affiliation, were set to missing. Thus, for each of the discrete variables

TABLE D-6

Case by Case Agreement between Actual Discrete Responses and Values Imputed Based on Other Variables

	Comparison with Data Imputed from Prior Year*		Comparison with Actual Data	
	Percentage Agreement between Imputed and Actual Responses	Estimated Percentage from Marginals***	Percentage Agreement between Imputed and Actual Responses	Estimated Percentage from Marginals***
Sex Served (3 categories)	87.1%	87.6%	88.1%	87.9%
Day or Boarding (3 categories)	88.2%	88.6%	84.3%	83.6%
School Type (4 categories)	38.5%	31.9%	46.6%	40.4%
Regular Curriculum (2 categories)	87.0%	86.9%	95.7%	95.1%
Special Education (2 categories)	75.3%	75.4%	92.5%	91.6%
Vocational Education (2 categories)	94.0%	94.0%	95.5%	95.5%
Compensatory Education (2 categories)	88.3%	88.4%	92.4%	92.4%
Foreign Language Served (2 categories)	88.0%	87.7%	95.3%	94.2%
Title I Support (2 categories)	79.1%	76.1%	57.7%	52.0%
Title IV B Support (2 categories)	55.5%	46.9%	78.3%	70.6%
Title IV C Support (2 categories)	90.5%	90.5%	77.0%	76.4%
School Lunch or Milk (2 categories)	49.7%	44.5%	60.4%	57.4%
Other Federal Assistance (2 categories)	81.1%	80.9%	83.0%	82.9%

*Based on all cases from the telephone (n=988) and nonrespondent (n=2805) where actual 1976-77 values were available but 1977-78 values were not.

**Based on all cases from the mail (n=1066) and telephone (n=988) samples where actual 1977-78 values were reported.

***These estimates are the rates of agreement that would occur if the actual and imputed marginals were matched but there were no further correlation between actual and imputed values.

imputed in the validation study, all of the predictor variables were themselves imputed, directly or indirectly, from religious affiliation. The low rate of agreement on a case-by-case basis does not pose any important problems for statistical analyses, since the bias in summary statistics is of primary concern in such analyses, but it does emphasize that imputed values should not be relied on in examining particular schools. In particular, it would seem wise not to include the imputed values on data tapes containing school identifiers for directory purposes.

Pass 3: Imputing Continuous Variables (Questions 12 through 15) from Prior Year's Values

For the continuous variables, changes from year to year (e.g., enrollment increases or declines) are likely to be important for addressing key issues policy. The [illegible]ation procedure of directly substituting the prior year's value would mask any such changes and so was rejected. A second possibility was to increase or decrease the prior year's value by an estimated overall growth rate for each variable. This was also rejected, since the differential growth rates for different types of schools is also an issue of concern. To reflect the major sources of variation in growth rates in the imputed values, the procedure selected for Pass 3 was to estimate the growth rate (expressed as the ratio of the 1977 value to the 1976 value) for each variable using a regression equation that incorporated the main discrete variables and those continuous variables already imputed.

To match the variances of the variables being imputed (as opposed to merely estimating the mean), it would have been necessary to add a random variable to the estimated growth rates which had a normal distribution with mean zero and variance equal to the residual variance after applying the regression equation. The resultant growth estimate (the regression function value plus the random component) would then be multiplied by the prior year's value to produce the imputed value.

During the initial validation of the Pass 3 procedures, it was discovered that the variances of the random component added to each regression value were too large. Most of the growth ratios were very near 1.0, but a

few schools with very large apparent growth rates added greatly to the variance of the growth rates and of the residual growth rates after the regression values were subtracted. A significant proportion of the large apparent growth rates were most probably the result of recording errors in one of the survey years. In addition, the growth rate residuals were highly skewed in a positive direction. The addition of a normally distributed random component did not introduce any mean bias but it did not reproduce this skewness well. As a result, the imputed values indicated a noticeably larger number of schools showing large decreases than did the reported values.

The addition of a random component to the regression values is essential to obtain accurate estimates of variances and correlations, but it also results in a greater variability (less accuracy) in the estimates of means or totals. Because of the strong emphasis on means and totals in the reporting of these data and because of the bias in the residual variances mentioned above, it was decided to reimpute the continuous variables using smaller variances for the random components. Values of one-tenth of the residual variances were tried, and because the results yielded more accurate mean estimates without substantially distorting estimates of variances, they were used in the final imputations.

Some of the continuous variables were breakdowns of totals (grade breakdowns of enrollment, sex breakdowns of graduates and staff, and categories of income and expenditures). In such cases, the total was estimated first and then allocated across the constituent variables in proportion to the prior year's allocation. This procedure was based on the observation that the allocations across the constituent variables were relatively stable across years. While this procedure simplified processing by greatly reducing the number of regressions that had to be performed, it may have masked somewhat any changes in the allocations across the constituent variables, so that researchers wishing to study such changes should be somewhat cautious in using the constituent values imputed by Pass 3.

One other problem with using a general linear model (regression) in this context was that the resultant estimates were occasionally negative. Because it was believed that the credibility of the data set and of the imputations would be damaged if negative values were left on the file, a method was instituted to remove the negative values while adding as little bias as possible. If the random component that was added to the estimated ratio of 1977 to 1976 values was larger in absolute value than regression function value, this would create either negative or very large estimates. In these cases, the random component was truncated by setting it equal, in absolute value, to the regression value. (For total enrollment, the random component was actually truncated at an absolute value two less than the regression value so that the resulting minimum would be two rather than zero.) Since this truncation was symmetric, no net bias was introduced. In a few cases, however, the regression value was itself negative (or less than two for total enrollment). In such cases, the regression value was raised to the minimum value. In order to avoid a net bias, an equivalent amount was subtracted from the next imputed value for the same variable.

We can express the imputation procedure for each variable, Y, as a series of equations. First, a regression estimate was obtained for each case, i., for which Y is missing.

$$\hat{R} = B_o + \sum_{j=1}^{p} \beta_j X_{ij} + \Delta$$

The value of Δ was either zero, or a value carried forth from a prior case. Next, two functions of the regression estimate were computed.

$$\sigma\left(\hat{R}_i\right) = \begin{cases} \hat{R}_i, & \text{if } \hat{R}_i \geq o \\ o, & \text{if } \hat{R}_i < o \end{cases}$$

$$\Delta\left(\hat{R}_i\right) = \begin{cases} o, & \text{if } \hat{R}_i \geq o \\ \hat{R}_i, & \text{if } \hat{R}_i < o \end{cases}$$

The value of Δ would be used on the next case requiring imputation of this variable. Third, a normal random variable, E^i, was computed with a zero mean and appropriate varance. Then a function of that error component was computed.

$$f(E_i) = \begin{cases} E_i, \text{ if } |E_i| \leq \hat{R}_i \\ E_i \left(\hat{R}_i / |E_i| \right), \text{ if } |E_i| > \hat{R}_i \end{cases}$$

Finally, the value to be imputed for Y was computed as a multiple of the prior year's value.

$$\hat{Y}_i = \left[g(\hat{R}_i) + f(E_i) \right] Y_i, \text{ prior year.}$$

One final correction was introduced to remove a bias in the imputed values caused by using growth ratios in the regression analyses rather than the final reported values. The extent of the bias was assessed by computing the mean of the inverse of the ratios as well as the mean of the ratios themselves. The extent to which the product of these means exceeded one was a measure of the potential bias introduced by the use of growth ratios as the basis for regression. This bias was appropriately eliminated by dividing each imputed value by the square root of the product of the mean ratio and the mean inverse ratio. Table D-7 shows the corrections that were applied to each value imputed in Pass 3.

In testing the Pass 3 imputations, two samples were drawn--a mail survey sample and a telephone survey sample. Since actual responses were needed for a criterion variable, it was not possible to test these procedures for nonrespondents, and the tests on the telephone sample were necessarily limited to the variables included in the telephone survey (i.e., total enrollment and classroom teachers).

Table D-8 shows the results of the validation of the Pass 3 procedures for the mail and telephone samples. The total enrollment was overestimated

TABLE D-7

Corrections to "Prior Year Estimates to Remove Bias in Growth Ratios

Variable	School Type	Mean* ratio of 1977 value to 1976 value	Mean* ratio of 1976 value to 1977 value	Correction**	Mean* ratio of 1978 value to 1977 value	Mean* ratio of 1977 value to 1978 value	Correction***
Total Enrollment	Elementary/Middle	1.004	1.031	1.017	1.028	1.018	1.023
Total Enrollment	Combined	1.133	1.002	1.065	1.070	1.045	1.057
Total Enrollment	High Schools	1.012	1.052	1.032	1.024	1.047	1.036
Total Enrollment	Other	1.075	1.242	1.156	1.194	1.052	1.121
Total Classroom Teachers	Elementary/Middle	1.048	1.018	1.033	1.034	1.033	1.033
Total Classroom Teachers	Combined	1.212	.963	1.080	1.094	1.047	1.070
Total Classroom Teachers	High Schools	1.075	1.010	1.042	1.051	1.021	1.036
Total Classroom Teachers	Other	1.336	1.067	1.194	1.292	1.064	1.173
Total Graduates	All	1.011	1.007	1.009	1.010	1.016	1.013
Total Religious Teachers	All	1.002	1.063	1.032	1.074	.998	1.035
Other Professional Staff	All	1.159	1.141	1.150	1.124	1.134	1.129
Other Staff	All	1.229	1.126	1.176	1.448	.967	1.183
Total Expenditures	All	1.199	.994	1.092	1.029	1.198	1.110
Total Income	All	1.263	1.131	1.195	1.092	1.295	1.189

* The mean is the mean of the ratios, not the ratio of the means.

** The first column of corrections was applied to 1977 values imputed from 1976 data and to 1976 values imputed from 1977 data. The correction is the square root of the product of the two means of ratios.

*** The second column of corrections was applied to 1977 values imputed from 1978 data and to 1978 values imputed from 1977 data.

TABLE D-8

Test of Imputations from Prior Year Data

Variable	N	Mean		Standard Deviation		Correlations with: No. of Students		No. of Teachers		Total Income		Correlation of Actual with Imputed Values
		Actual	Imputed	Actual	Imputed	Actual	Imputed	Actual	Imputed	Actual	Imputed	
Mail Survey Sample												
Total Number of Students	1029	302.2	301.6	236.8	236.1	x	x	.82	.78	.55	.41	.97
Number of Preschoolers	910	3.5	3.3	11.9	11.9	.04	.07	.14	.12	.1	.11	.85
Number of 1st Graders	910	26.5	26.5	22.2	22.9	.44	.46	.16	.16	.00	.00	.92
Number of 4th Graders	910	25.7	25.5	22.2	22.5	.47	.48	.17	.15	.01	.02	.94
Number of 6th Graders	910	26.8	27.5	24.0	24.9	.48	.49	.19	.17	.05	.00	.94
Number of 8th Graders	910	28.0	28.2	27.0	28.0	.50	.53	.28	.29	.19	.16	.95
Number of 9th Graders	910	20.6	20.1	60.6	58.3	.65	.62	.68	.65	.55	.41	.99
Number of 12th Graders	910	17.6	17.2	51.0	49.6	.65	.63	.69	.67	.57	.44	.98
Number of Graduates	995	16.4	14.5	48.3	42.9	.62	.60	.68	.64	.55	.42	.98
Number of Classroom Teachers	978	15.0	15.2	12.5	13.1	.82	.78	x	x	.85	.74	.91
Number of Male Teachers	932	3.5	3.7	7.8	7.2	.55	.53	.72	.75	.76	.69	.95
Number of Other Professionals	818	2.9	3.2	3.6	3.9	.32	.34	.58	.60	.72	.66	.82
Number of Other Male Professionals	814	.9	1.0	2.2	2.4	.23	.26	.48	.52	.66	.68	.85
Number of Other Staff	826	5.1	6.4	6.1	8.2	.45	.37	.66	.61	.75	.71	.83
Number of Male Other Staff	821	1.5	2.0	2.4	3.3	.34	.27	.53	.50	.65	.66	.74
Total Expenditures*	735	258.1	290.0	357.9	484.5	.58	.43	.87	.76	.98	.94	.90
Current Operations	581	261.6	292.3	362.9	469.6	.60	.46	.90	.81	.97	.95	.91
Total Income*	720	254.1	255.7	355.8	413.3	.55	.41	.85	.74	x	x	.86
Current Operations	555	252.0	258.0	350.0	399.2	.55	.41	85	.78	.98	.99	.85
Telephone Survey Sample												
Number of Students	497	242.7	254.5	256.6	281.4	x	x	.75	.86	-	-	.94
Number of Teachers	379	15.1	16.4	13.9	16.1	.75	.86	x	x	-	-	.80

* Expenditure figures are in thousands of dollars.

slightly in the telephone sample and underestimated very slightly in the mail sample. The two samples did differ significantly in terms of their mean enrollment (243 vs. 302) and roughly 80% of this difference is reflected in the imputed means (255 vs. 302). The remainder reflects differences between the two samples that were not captured in the variables used to predict total enrollment. All of the other means also show a very close correspondence to the actual values, although the expenditure imputations in the mail sample are noticeably higher than the actual values.

The estimates of the standard deviations are also quite close to the actual values. The standard deviation of total enrollment was overestimated slightly in the telephone sample and underestimated very slightly in the mail sample as was the mean. The standard deviations of the income and expenditure data were overestimated somewhat in this sample of mail respondents. Overall, however, the approximation is quite good, even though the error variances for the imputed values had been divided by 10.

The third set of numbers in Table D-8 are the actual and imputed intercorrelations among key variables. The correlations among the numbers of students, teachers, and graduates were reproduced with a high degree of accuracy. Some of the correlations for "other staff" and for the budgetary data show consistent attenuation. The source of the attenuation for these particular variables was at least in large part the presence of many predictors in the regression equations for these variables, because they were imputed last using the results of imputations of other variables. As a consequence, the relationship between each of these variables and any single predictor played a relatively smaller part in the total prediction than it would have if there were fewer predictors. Another source of the problem, however, was that the budgetary variables and to a certain extent the staffing variables are very highly skewed so that correlations with these variables tend to be dominated by a relatively small number of cases with very large values. In addition, the overestimate of the standard deviations for the budgetary variables is related directly to underestimates of associated correlations.

The final set of numbers in Table D-8 are the correlations between the actual and imputed values for each case. These correlations are quite high, ranging from .74 up to .99, reflecting the consistency between the current and prior year's values. The values are sufficiently high, in fact, that the imputed values for a particular case could be used on a case-by-case basis for many purposes (e.g., in drawing samples).

Pass 4: Imputing Continuous Variables from Current Year Variables Only

The final set of imputation procedures used with the Nonpublic School Survey were applied to cases where continuous variables were missing and prior year values also were not available. The general procedure was to regress the variable to be imputed on a set of recoded discrete variables and previously imputed continuous variables. The variables actually estimated by regression, however, were generally not the variables on the file but ratios of the variables on file to other variables (e.g., student/teacher ratio, per-pupil expenditures). This approach was taken because the ratios had relatively smaller coefficients of variation than the variables themselves and hence could be predicted with greater sensitivity.* Table D-9 shows the variables actually used in the imputations and the multiple correlation coefficients indicating the accuracy with which each was predicted.

The variables were imputed in the order shown in Table D-9, and each regression dependent variable (including its random error component) was used as a predictor in subsequent regressions. All regressions included religious affiliations, sex of students served, day-boarding status, and community size. The intermediate variables were selected to have much smaller coefficients of variation than the target dependent variables, to allow the regression analyses to be more sensitive. If an intermediate variable could be found with a coefficient of variation nearly equal to zero, then the imputation would be nearly perfect even if the multiple correlation for the regression were quite small. For example, although

*For example, whether the school was a boarding school or not is strongly related to per pupil expenditures, but the relationship to raw expenditures would be swamped by the effect of school size.

TABLE D-9

Order of Imputation of 1977-78 Continuous Variables from Other 1977-78 Data, Intermediate Variables, Coefficients of Variation, and Multiple Correlations for Intermediate Variable Predictions

Variables (Y)	Intermediate Variables (Regression Dependent Variables) (Z)	CV_Y*	CV_Z	Multiple Correlation**
Total Enrollment	(none)	.9	-	.5
Total Classroom Teachers	(Teacher/Pupil) Ratio	1.0	.8	.7
Grade Enrollment Breakdown	Lowest Grade	3.2 (for 12th graders)	1.0	.9
	Highest Grade		- .2	.8
Total Graduates	(Graduates/Twelfth Grade Enrollment) Ratio	3.4	.1	.2
Total Expenditures	Expenditures Per Pupil	2.4	1.7	.7
Total Income	(Income/Expenditure) Ratio	4.2	.5	.2
Male Classroom Teachers	(Male/Total) Classroom Teachers	2.2	1.1	.9
Male Graduates	(Male/Total) Graduates	3.7	.2	.9
Total Religious Teachers	Religious/Total Classroom Teachers	1.8	.9	.3
Total Other Professionals	(Other Professional/Classroom Teacher) Ratio	1.6	1.5	.5
Total Other Staff	(Other Staff/Classroom Teacher) Ratio	3.1	1.2	.6
Male Religious Teachers	(Male/Total) Religious Teachers	5.4	1.2	.8
Male Other Professionals	(Male/Total) Other Professionals	2.5	1.3	.6
Male Other Staff	(Male/Total) Other Staff	4.3	.9	.4
Current Operating Expenditures	(Current/Total) Expenditures	2.3	.1	.7
Current Operating Income	(Current/Total) Income	4.4	.1	.8
Tuition Income	(Tuition/Total) Income	5.0	.6	.6
Tuition for Current Operations	(Tuition for Current Operations/Tuition Total) Income	5.2	.1	.6

*Coefficient of variation = standard deviation/mean.

**Multiple correlations of combinations of regression predictor variables for predicting variation in the regression dependent variable z. The predictive power (correlation) of the mean value of z for predicting y is approximately $\sqrt{1 - CV_z^2/CV_y^2}$.

the number of graduates varied greatly across schools (CV=3.4), the number of graduates was closely related to the number of twelfth graders (CV_{ratio}=0.1), so using the mean value for the ratio would provide reasonably accurate estimates even though other 1977-78 data only accounted for a small proportion ($.24^2$ = .06) of the variance of the ratio.

In Pass 4 as in Pass 3, a random error component was added to the regression estimate in order to avoid underestimating variances, and the same error truncation procedure was used to truncate both the error component and the regression estimate itself so as to avoid negative values without introducing a mean bias. Pass 4 did not, however, include any correction for bias introduced by ratio estimators. Total enrollment was estimated directly and so was not subject to the bias. Other estimates did not show any appreciable bias since the denominator was constrained to be greater than the numerator (e.g., % male teachers).

For a few variables, a more noticeable bias was introduced. The imputed numbers of teachers, for example, were approximately 10% higher than actual values in the validation test. The problem of maintaining agreement between total and constituent variables was solved by first imputing the total variable and then imputing the ratio of the first constituent variable to the total under the constraint that this ratio lie between 0 and 1. The second constituent variable was then computed by subtraction. The only case where there were more than two constituent variables for a given total, school enrollment, was treated in a much different manner. It was recognized that the grade enrollments were partially discrete in that schools would necessarily have a value of 0 for the grades not served. Consequently, the range of grades served was imputed first and then the allocation of the total enrollment across those grades was imputed. The two key variables, "lowest grade served" and "highest grade served," were defined by considering preschoolers to be one grade lower than kindergarten and the postgraduate, special, and ungraded categories to follow 12th grade. The allocation of the total enrollment across grades served was made either in proportion to the average grade distributions for schools with exactly the same range of grades served

or, if there were not sufficient cases reporting grade enrollments for that particular grade range, equally across grades in the range.

In testing the Pass 4 imputations, three samples were drawn corresponding to the three samples drawn for the Pass 2 validation. For the telephone survey cases, two samples were actually drawn, a general sample for the variables included in the telephone survey and a sample with prior year's data for the variables not included in the telephone survey. The non-respondent sample was also limited to cases with prior year's data with the Pass 3 imputations being used as the criterion. For each validation sample, the actual values and the Pass 3 imputations were masked and new values were imputed using Pass 4. (For the second telephone sample, however, the actual telephone survey variables were not masked and were not reimputed to match better the availability of this information when 4 was used to impute non-telephone survey variables on the telephone survey sample.)

Tables D-10, D-11, and D-12 show the comparisons of the actual and imputed values for this test for the mail respondents, telephone respondents, and nonresponding prior year respondents, respectively. The imputation of student enrollments appears to have been reasonably successful. The imputed mean was slightly lower than the actual mean for the mail survey sample and somewhat higher for the telephone and nonrespondent samples. The imputed mean for the three groups as a whole extrapolated to the entire file, (256.2) matches the actual mean (250.3) quite closely and, in addition, more than two-thirds of the variation between the three samples (291.8, 172.6, and 181.4) is reflected in the imputed sample means (283.0, 202.5, and 215.7).

The variance estimated was noticeably lower for all three samples. This was the result of truncation of the random component to yield more accurate mean estimates and to avoid negative values. While the bias in the variance estimates appears to be substantial (roughly 25%) it must be remembered that only 5% of the schools had enrollments imputed in Pass 4.

TABLE D-10

Test of "Current Year Only" Imputations for Continuous Variables (Mail Survey)*

Variable	N	Mean		Standard Deviation		Correlations with: No. of Students		Correlations with: No. of Teachers		Correlations with: Total Income		Correlation of Actual with Imputed Values
		Actual	Imputed	Actual	Imputed	Actual	Imputed	Actual	Imputed	Actual	Imputed	
Total Number of Students	1066	291.8	283.0	236.3	162.2	x	x	.82	.66	.55	.52	.46
Number of Preschoolers	1021	3.6	1.9	11.7	8.4	-.01	.02	.05	.06	.05	.10	.04
Number of 1st Graders	1021	25.6	25.4	21.4	17.3	.39	.22	.11	-.13	-.12	-.24	.55
Number of 4th Graders	1021	24.8	25.3	21.7	16.5	.41	.24	.14	-.11	-.09	-.24	.55
Number of 6th Graders	1021	25.8	25.7	23.2	17.2	.44	.26	.18	-.09	-.05	-.21	.54
Number of 8th Graders	1021	25.7	20.9	26.0	21.0	.46	.36	.26	.16	.10	.08	.31
Number of 9th Graders	1021	20.3	17.6	59.1	46.4	.69	.66	.71	.63	.59	.63	.68
Number of 12th Graders	1021	17.6	16.7	52.5	43.8	.69	.68	.71	.65	.60	.62	.73
Number of Graduates	1035	19.0	17.3	64.6	45.2	.61	.67	.62	.67	.49	.61	.66
Number of Classroom Teachers	1066	14.3	15.3	11.7	11.8	.82	.66	x	x	.81	.85	.48
Number of Male Teachers	1066	3.7	4.1	6.6	7.7	.50	.49	.70	.76	.77	.78	.66
Number of Other Professionals	1066	2.5	2.9	3.0	3.3	.39	.32	.55	.70	.64	.82	.49
Number of Other Male Professionals	1066	.8	.7	1.6	1.6	.31	.23	.50	.59	.68	.81	.58
Number of Other Staff	1066	4.8	5.1	7.4	5.9	.33	.35	.55	.70	.75	.88	.41
Number of Male Other Staff	1066	1.6	1.8	3.5	2.2	.22	.35	.43	.61	.69	.83	.41
Total Expenditures**	873	244.6	249.5	332.9	272.4	.55	.52	.82	.86	.99	.99	.54
Current Operations	859	228.2	228.3	298.9	246.0	.57	.54	.83	.87	.98	.99	.52
Total Income**	849	241.7	239.6	335.0	278.0	.55	.52	.81	.85	x	x	.53
Current Operations	811	224.1	220.7	292.9	236.6	.55	.54	.82	.87	.98	.997	.55

* Based on a sample consisting of every 12th mail survey respondent.

** Income and Expenditure figures are in thousands of dollars.

TABLE D-11

Test of "Current Year Only" Imputations for Continuous Variables (Telephone Survey)*

Variable	N	Mean		Standard Deviation		Correlations with: No. of Students		No. of Teachers		Total Income		Correlation of Actual with Imputed Values
		Actual	Imputed	Actual	Imputed	Actual	Imputed	tual	Imputed	Actual	Imputed	
Total Number of Students	956	172.6	202.5	196.3	145.7	x	x	.78	.84	-	-	.29
Number of Preschoolers	501	9.2	4.9	25.9	14.3	-	-	.20	.16	.06	.15	.37
Number of 1st Graders	501	20.6	18.9	22.2	19.3	-	-	.05	-.07	.00	.04	.85
Number of 4th Graders	501	16.0	18.4	17.7	17.5	-	-	.04	-.06	.03	.11	.85
Number of 6th Graders	501	16.0	18.5	19.0	18.5	-	-	-.01	-.03	.09	.19	.80
Number of 8th Graders	501	16.6	15.3	22.3	21.6	-	-	.26	.13	.16	.34	.70
Number of 9th Graders	501	17.2	17.0	39.1	35.5	-	-	.61	.61	.37	.55	.92
Number of 12th Graders	501	12.5	13.7	30.0	31.8	-	-	.60	.62	.39	.55	.91
Number of Graduates	199	29.2	34.7	35.7	44.3	-	-	-	-	-	-	.84
Number of Classroom Teachers	772	11.3	17.1	11.1	13.1	.78	.84	x	x	-	-	.29
Number of Male Teachers	772	3.0	5.3	4.5	6.1	.58	.59	.74	.77	-	-	.37
Number of Other Professionals	562	3.4	3.3	5.2	3.2	-	-	.55	.84	.65	.44	.51
Number of Other Male Professionals	557	1.4	1.1	2.4	1.4	-	-	.55	.79	.63	.45	.55
Number of Other Staff	562	6.8	5.5	16.4	5.9	-	-	.20	.78	.61	.52	.50
Number of Male Other Staff	554	1.9	1.6	5.2	1.7	-	-	.17	.68	.65	.46	.39
Total Expenditures**	271	369 8	358.2	385 3	585.9	-	-	.68	.88	.95	.98	.75
Current Operations	237	358.0	357.7	414.1	594.5	-	-	.80	.89	.92	.97	.75
Total Income**	254	296.9	324.7	486.9	483.6	-	-	.73	.87	x	x	.69
Current Operations	222	278.0	319.6	522.7	488.0	-	-	.82	.89	.97	.997	.68

* A general sample consisting of every fourth telephone respondent was used for those variables included in the telephone survey (total number of students and number of teachers). For the other variables a separate sample of telephone respondents with prior year's data was used and the imputations from the prior year's data were used as the "actual" values.

** Income and Expenditure figures are in thousands of dollars.

TABLE D-12

Test of "Current Year Only" Imputations for Continuous Variables (Nonrespondents)*

Variable	N	Mean		Standard Deviation		Correlations with: No. of Students		No. of Teachers		Total Income		Correlation of Actual with Imputed Values
		Actual	Imputed	Actual	Imputed	Actual	Imputed	Actual	Imputed	Actual	Imputed	
Total Number of Students	2368	181.4	215.7	214.1	163.7	x	x	.83	.75	.37	.47	.42
Number of Preschoolers	1552	5.4	3.2	17.5	9.6	.06	.03	.10	.01	.01	-.01	.02
Number of 1st Graders	1552	15.5	18.5	20.5	16.6	.49	.36	.23	-.03	-.01	-.26	.50
Number of 4th Graders	1552	13.7	18.2	18.6	15.8	.52	.39	.25	.00	.00	-.25	.54
Number of 6th Graders	1552	14.5	18.2	20.1	16.5	.51	.42	.25	.04	.02	-.23	.56
Number of 8th Graders	1552	15.1	14.7	23.4	20.1	.52	.53	.32	.28	.05	.07	.34
Number of 9th Graders	1552	15.4	18.4	48.1	43.0	.71	.70	.63	.70	.34	.63	.66
Number of 12th Graders	1552	12.2	16.1	42.0	39.3	.70	.68	.63	.69	.38	.63	.61
Number of Graduates	348	53.8	63.4	88.3	62.6	.81	.89	.55	.65	.23	.45	.39
Number of Classroom Teachers	2307	11.1	14.9	12.7	12.3	.83	.75	x	x	.67	.76	.29
Number of Male Teachers	1612	3.5	5.0	6.5	8.7	.55	.53	.73	.78	.56	.80	.45
Number of Other Professionals	1212	3.3	3.8	6.0	5.3	.28	.21	.54	.51	.48	.83	.24
Number of Other Male Professionals	1197	1.3	1.3	2.8	2.7	.24	.17	.48	.45	.47	.84	.34
Number of Other Staff	1231	5.9	6.1	12.4	8.7	.26	.32	.52	.61	.61	.88	.36
Number of Male Other Staff	1217	1.9	2.1	5.5	3.2	.20	.34	.47	.60	.64	.87	.34
Total Expenditures**	871	284.4	295.1	576.1	351.3	.37	.52	.67	.84	.96	995	.28
Current Operations	745	243.7	271.4	478.1	296.7	.42	.54	.76	.87	.90	.991	.25
Total Income**	818	234.6	288.2	412.9	361.5	.37	.47	.61	.76	x	x	.26
Current Operations	701	213.9	263.2	370.5	291.5	.41	.47	.66	.79	.94	.996	.24

* The test was run on a 50% sample of 1977 nonrespondents for which 1976 data were available. The "actual" values were inferred from the prior year's values using the imputation from prior year procedure.

** Income and Expenditure figures are in thousands of dollars.

The grade breakdowns of enrollment showed a reasonably good approximation to the actual values for the middle range of grades. Because the range of grades served was not linearly related to the predictor variables (largely because of the treatment of the preschool and special and ungraded categories) the prediction was less accurate for the extreme ranges of the grade breakdowns. The fact that the standard deviations also match reasonably well, notwithstanding the highly skewed nature of the variables due to the large number of zeroes for grades not served, demonstrates the value of the use of low and high grades as intermediate variables. The imputed correlations between total enrollment and the enrollments for specific grades approximate the patterns of actual correlations quite closely although there was some attenuation in the lower grades. In addition, except for the preschool enrollment, the imputed grade enrollments are significantly correlated with the actual values in all three samples.

The imputed means for the number of teachers were higher than the actual means for all three samples. This positive bias can be directly attributed to the use of a "ratio" variable, teachers per student, in the regression analyses and the failure to correct for the bias in the actual variable associated with the use of this ratio variable. Subsequent analyses indicated that this bias was on the order of 12% for cases imputed in Pass 4. If the imputed estimates were reduced by this amount, the imputed teacher means would show the same pattern as the enrollment estimates across the three samples.

The imputed and actual standard deviations for the number of teachers were very similar (11.8 versus 11.7; 13.1 versus 11.1; and 12.3 versus 12.7). Any negative bias due to the truncation of the random component was apparently balanced by a positive trend due to the use of the ratio estimator. The pattern of correlations between the number of teachers and other continuous variables was reproduced quite closely although the correlations with other staff counts were noticeably inflated in some instances. This last effect resulted from the truncation of the random component in the estimate of the ratio of other counts to the number of teachers. As a result of this truncation, the other counts were more nearly linear functions of the number of teachers in the imputed values than they were in the actual values.

The imputed means for the other staff counts were quite close to the actual values, with a slight trend toward overestimation reflecting the tendency toward overestimation in the teacher counts. The imputed standard deviations of the other counts were noticeably low relative to the actual standard deviations. This was the result of a much greater truncation of the random component in the imputations of these values. The standard deviation of the random component was set to one-half the regression error standard deviation in estimating enrollments, grade breakdown, and total teachers and to one-tenth the regression error standard deviation for all other Pass 4 imputations. While this truncation led to reduced standard deviations, it also increased the correlations between actual and imputed values, which as a result were highly significant in all three samples.

The budgetary data were reproduced quite well in most instances. (The budgetary variables were both highly skewed and relatively independent of other variables so that obtaining estimates of the correct order of magnitude is no small feat.) For the income data in the telephone sample and both the income and expenditure data in the nonrespondent sample, there were overestimates that could be associated with the use of ratio variables in the regression analyses, but the other imputations were surprisingly accurate. The standard deviations were overestimated in some instances, underestimated in others, and quite accurate in still others. Because of the highly skewed nature of these variables, the standard deviations were dominated by a very small number of schools in each instance, so the lack of consistency in these estimates should not be surprising.

The imputed correlations between total income and other variables were also dominated by a relatively small number of cases but were, nonetheless, quite close to the pattern of actual values. The tendency to overestimate was the direct result of the truncation of the component in the imputed values. Had a larger random component been added, lower correlations would have been obtained. (It is possible that comparison of the correlations would give a new method for estimating how large the random component should be.)

The final imputation step was the imputation of open-closed status among the nonrespondents. The open-closed status was not known for 3,473 nonrespondents in 1976, for 386 in 1977, and for 529 in 1978. A sample of 114 of the first group was contacted, and 8 were found to be closed; 13 of 83 in the second group were closed; and 3 of 19 in the last group were closed. These base rates reflect a complex interaction between school openings and closing and survey response follow-up methodology. Each missing open-closed status was imputed randomly, separately for schools with small (< 75), medium, and large (> 175) enrollments, matching the base rates from the telphone contacts. For small, medium, and large schools, these "closed" rates were .096, .064, and .021 for 1976; .381, .212, and .058 for 1977; and .326, .127, and .037 for 1978. The corresponding numbers of schools requiring imputation of open-closed status were 1,758, 1,074, and 641 for 1976; 72, 96, and 218 in 1977; and 229, 128, and 172 in 1978.

Summary

The imputation of data in the nonpublic school survey was undertaken in an exploratory mode, to find out what could be done while responding to a real need for information on the nonpublic school population. The availability of three years' data was an essential component of the methods used because, for the most part, responses from a single school would not be expected to change dramatically except in a small number of cases. The purpose of the imputation was to provide a file that could be used by future analysts to compute arbitrary summary statistics, not to provide the basis for identifying particular schools. The individual imputed values are, of course, subject to error; and future data collection should be carried out with the perspective of reimputing where appropriate. In particular, schools imputed to have closed should be included in sampling frames as if they were open. Revision of imputed values based on new data collection will introduce bias, however and this must be investigated in the future.

Given that a number of differences between imputed and actual means were discovered, it is important to determine the practical implications of these differences for use of the file. Tables D-13 and D-14 show the

TABLE D-13

Number of Cases with Real and Imputed Discrete Responses, by Respondent Population

Variable	Actual Data				"Prior Year" Method*				"Current Year Only"			
	Mail	Telephone	N/R	Total	Mail	Telephone	N/R	Total	Mail	Telephone	N/R	Total
Population Center	12833	3953	0	16786	15	7	2496	2518	1	2	768	771
Sex Serve	12833	3932	0	16765	15	29	2494	2538	1	1	770	772
Day or Boarding	12833	3929	0	16762	15	32	2494	2541	1	1	770	772
School Type	12833	3950	316	17099	15	11	2494	2520	1	1	454	456
Curriculum	12828	3	0	12831	16	2041	1682	3739	5	1918	1582	3505
Foreign Language	12828	3	0	12831	16	2041	1682	3739	5	1918	1582	3505
Federal Assistance												
Title I	12816	2	0	12818	28	2042	1682	3752	5	1918	1582	3505
Title IV-B	12823	2	0	12825	21	2042	1682	3745	5	1918	1582	3505
Title IV-C	12806	2	0	12808	38	2042	1682	3762	5	1918	1582	3505
Lunch	12825	2	0	12827	19	2042	1682	3743	5	1918	1582	3505
Other	12792	2	0	12794	52	2042	1682	3776	5	1918	1582	3505

TABLE D-14

Number of Cases with Real and Imputed Continuous Responses, by Respondent Subpopulation

Variable	Actual Data				"Prior Year" Method*				"Current Year Only"			
	Mail	Telephone	N/R	Total	Mail	Telephone	N/R	Total	Mail	Telephone	N/R	Total
Total Enrollment	12829	3841	0	16670	19	69	2323	2411	1	52	941	994
Enrollment by Grade	12829	2	0	12831	19	1973	2362	4354	1	1987	902	2890
Total Teachers	12824	3351	0	16175	22	574	2314	2910	3	37	950	990
Teachers by Sex	12824	3101	0	15925	22	551	1725	2298	3	310	1539	1852
Religous Teachers	12248	3579	2442	18869	1	360	771	1132	0	23	51	74
Religious Teachers by Sex	12818	3363	2442	18653	1	380	331	712	0	219	491	710
Other Professionals	12825	0	0	12825	17	1280	1218	2515	7	2682	2046	4735
Other Professionals by Sex	12825	0	0	12825	19	1367	1284	2670	5	2595	1980	4580
Other Staff	12825	0	0	12825	16	1190	1170	2376	8	2772	2094	4874
Other Staff by Sex	12825	0	0	12825	21	1329	1255	2615	3	2633	1999	4635
Total Graduates												
Graduates by Sex	11909	2143	2154	16206	57	33	122	212	883	1786	988	3657
Budget												
Tuition for Curr. Ops.	10111	1	0	10112	202	402	95	699	2536	3559	3169	9264
Total Tuition	10497	1	0	10498	210	402	95	707	2142	3559	3169	8870
Total for Curr. Ops.	9784	1	0	9785	976	910	819	2705	2089	3051	2445	7585
Total Income	10769	1	0	10770	1229	1139	1104	3472	851	2822	2160	5833
Expenditures for Curr. Ops.	10318	1	0	10319	1145	1042	960	3147	1386	2919	2304	6609
Total Expenditures	10513	1	0	10514	1423	1146	1114	3683	913	2815	2159	5878

* From 1976 or 1978

frequency with which each variable was imputed from either the prior year or other variables in the current year for 1977. From these frequencies it is possible to derive estimates of the bias introduced in overall summary statistics from bias produced during imputation. Furthermore, from these frequencies and from estimates of the accuracy of imputations it will be possible to estimate the error of measurement introduced by missing data and the reduction in that error through imputation.

Tables D-15 and D-16 show the "bottom line," the degree of error in key statistics for the total file, estimated from the findings of the validation study. The degree of bias estimated for mail respondents, telephone respondents, and nonrespondents from each imputation pass is taken from Tables D-3 and D-4 for discrete variables (Table D-15) and from Tables D-10, D-11, and D-12 for the continuous variables (Table D-16). For each category, the bias is estimated by the difference between the imputed mean and the actual mean for the appropriate validation sample. In Table D-15, the bias has been expressed as a percentage of the actual value. An average bias due to missing data for the total sample is estimated by assigning the estimated bias for each method to each value imputed with that method and averaging across all schools including those with reported values (for which the bias is zero).

For the discrete variables, the estimated total bias is comfortably low, less than one-half of one percent for the variables analyzed. The estimated bias for the continuous variables is somewhat larger, but is still small relative to the probable bias due to incomplete coverage of the private school universe. (One investigation of several states indicated that as many as 5% to 10% of the universe may be missing from the file.) The larger amounts of bias in the teacher and expenditure data reflected, in part, failure to correct for the use of ratio variables in the regression analyses. It should also be noted that the bias estimates in Table D-16 reflect the assumption that the "bias" in the criterion used for the Pass 4 "nonrespondent sample" was the same as the bias estimated from the Pass 3 test sample. (This caused the final estimated bias to be larger than the observed difference since the two biases were added.)

TABLE D-15

Estimated Percentage Bias in Reported Totals for Selected Discrete Variables

	Variable			
	No. of Schools Categorized as "Combined"	No. of Schools Categorized as "Day Only"	No. of Schools Offering a Special Education Curriculum	No. of Schools Receiving ESEA Title I Support
Mail Survey Respondents, Variables Imputed from Prior Year	+0.0% (n=15)	+0.3% (n=15)	-0.1% (n=16)	-2.1% (n=21)
Telephone Survey Respondents, Variables Imputed from Prior Year	* (n=32)	* (n=32)	* (n=2041)	* (n=2042)
Nonrespondents, Variables Imputed from Prior Year	* (n=2494)	* (n=2494)	* (n=1682)	* (n=1682)
Mail Survey Respondents, Variables Imputed from Other Variables	+2.2% (n=1)	-0.6% (n=1)	-0.2% (n=5)	-0.6% (n=5)
Telephone Survey Respondents, Variables Imputed from Other Variables	-4.6% (n=1)	-1.5% (n=1)	* (n=1918)	* (n=1918)
Nonrespondents, Variables Imputed from Other Variables	* (n=770)	* (n=770)	* (n=1582)	* (n=1582)
Total Percentage Bias, including Cases with Actual Data**	-0.2% (n=20075)	-0.0% (n=20075)	-0.1% (n=20075)	-0.5% (n=20075)
Reported Total Estimate, after Imputation Imputation	3451	18357	1531	6351
Revised Total Estimate After Imputation Bias Correction	3486	18361	1542	6451
(Estimate if Nonrespondents Deleted from Universe)	(2704)	(15636)	(631)	(5008)
(Estimate if Nonrespondents Assigned Mean Value for Respondents)	(3174)	(18726)	(987)	(7875)

* No estimate could be obtained for these cells. The percentage bias is estimated to be the same as in the cell immediately above for the purpose of developing a total estimate.

** The total bias was estimated by averaging the biases above and a bias value of zero for schools with a reported value. The averages are weighted by the number of schools to which each bias estimate applies.

TABLE D-16

Estimated Percentage Bias in Reported Totals
for Key Continuous Variables

Category	Variable		
	Enrollment	Classroom Teachers	Current Operating Expenditures (in dollars)
Mail Survey Respondents, Variables Imputed from Prior Year	-.2% (n=19)	1.3% (n=22)	11.7% (n=1145)
Telephone Survey Respondents, Variables Imputed from Prior Year	4.9% (n=69)	8.6% (n=574)	* (n=1042)
Nonrespondents, Variables Imputed from Prior Year	* (n=322)	* (n=2184)	* (n=960)
Mail Survey Respondents, Variables Imputed from Other Variables	-3.0% (n=1)	7.0% (n=3)	0.0% (n=1386)
Telephone Survey Respondents, Variables Imputed from Other Variables	17.3% (n=52)	51.3% (n=37)	11.6%** (n=2919)
Nonrespondents, Variables Imputed from Other Variables	24.7%** (n=941)	45.8%** (n=950)	24.4%** (n=2304)
Total Percentage Bias, including Cases with Actual Data	1.4% (n=20075)	3.1% (n=20075)	6.3% (n=20075)
Reported Total Estimate, after Imputation	5,140,000	278,161	5,103,000,000
Revised Estimate, after Imputation Bias Correction	5,067,000	269,509	4,783,000,000
(Estimate if Nonrespondents Deleted from Universe)	(4,522,000)	(226,647)	(2,633,000,000)
(Estimate if Nonrespondents Assigned Mean Value for Respondents)	(5,449,000)	(285,710)	(5,122,000,000)

* No estimate could be obtained for these cells. The percentage bias is estimated to be the same as in the cell immediately above for the purpose of developing a total estimate.

** This bias is based on a comparison between imputations using the two different methods, adjusted for the bias estimated for the imputation from prior year's data. In all other cases the bias is estimated by comparing imputed to actual responses.

These results should also be taken as tentative because of other assumptions about similarities between the true nonrespondents and the nonrespondent and telephone samples used in these analyses. Still, the bias estimates in Tables D-15 and D-16 are clearly the best information available. Finally, it is important to note that these biases do not pose serious problems for comparisons between years (i.e., growth estimates) because the 1976 and 1978 values were always imputed from the corresponding 1977 value using Pass 1 and 3 procedures, for which little bias was found. Parenthetically, estimates based on two simple alternatives to the procedures actually used are shown at the bottom of Tables D-15 and D-16.

The results of this study also had important implications for the development and use of missing data algorithms. Most particularly, the general regression approach, which is the procedure of choice in some available imputation software (e.g., BMDPAM), performed poorly for the continuous variables in this study. This failure resulted from the facts that (1) some of the relationships among the variables were not linear (2) the conditional distributions (conditional on the regression function value) were non-normal and highly skewed, and (3) the variances of the conditional distributions were not at all constant. In a related SAGE task, a procedure is being developed that overcomes these difficulties by splitting the variable to be imputed into a number of discrete levels and then using an approach similar to the discrete variable approach used in Pass 2. Under this procedure, separate conditional distributions are estimated for different levels of the regression function values with no assumptions regarding the linearity of the means or homogeneity of the variances of these distributions. The very positive results for the Pass 2 procedure are highly encouraging with respect to this new procedure.