CCD Data File:
Thirteen-Year Longitudinal Common Core of Data Non-Fiscal Survey Database:
School Years 1986-87 – 1998-99

Don McLaughlin
American Institutes for Research

June 11, 2003

Table of Contents

List of Tables

Overview

This report describes (a) the procedures used to continue the development of the thirteen-year Longitudinal Common Core of Data (CCD) Non-Fiscal Survey Database and (b) the files resulting from that effort. The processes carried out during the period from September 2001 to April 2003 resulted in the addition of 1998-99 data to the longitudinal database, including district-level measures, school-level measures, and district staff categories. These files are designed for research use in testing hypotheses about longitudinal trends in schools and school districts over the period from 1986-87 to 1998-99. To facilitate analysis, all missing data on selected measures have been replaced by statistical imputations, and clearly erroneous responses have been edited and replaced by plausible values.

The file development procedures described in this report represent the fifth round of CCD editing and imputation undertaken by the author. The editing and imputation procedures used in this round were adapted from the previous rounds of editing and imputation, described in Appendix B of this report. The most notable changes in the Longitudinal Database resulting from addition of the 1998-99 data are:

(1)   Renaming of all variables to match current NCES variable names;

(2)   Addition of gender and race breakdowns for diploma recipients and other high school completers;

(3)   Addition of Bureau of Indian Affairs (BIA) districts and schools to the file;

(4)   Addition of counts of students with limited English proficiency (LEP); and

(5)   Replacement of a single count of free-lunch eligible students with separate counts of free-lunch eligible students, reduced price lunch eligible students, and the sum of the two counts.

In the first round, missing data in six years were imputed to provide the basis for a district-level longitudinal trend report based on the years 1986-87 through 1991-92 (Levine, McLaughlin, and Sietsema, 1995). In that round, no test for outliers was carried out, and no reported values were replaced with more consistent imputed values. It became apparent during the analyses for that report that some reported values were probably in error. Therefore, in the second round, tests for outliers were incorporated into the file development procedure; and 1992-93 and 1993-94 data were added to the file to provide the basis for a longitudinal trend report on small rural school districts (McLaughlin, Huberman, Hawkins, and Hoffman, 1997). The imputation procedures for the second and later rounds are described in that report and are included in Appendix B of this report. The third round included addition of 1994-95 and 1995-96 data to the Longitudinal District File, identification of linkages across years for districts that consolidated and between grade levels for separate elementary and secondary districts, and addition of the five outlying territories to the file. The fourth imputation round included adding the 1996-97 and 1997-98 data to the Longitudinal District File; addition of a Longitudinal School File, with school data for 1986-87 through 1997-98, to the database; and addition of staffing by category counts for the years 1992-93 through 1997-98 to the database.

The Longitudinal Database described in this report is derived from the basic CCD data collected by NCES.[1] The basic CCD Local Agency Nonfiscal File for each year contains records for roughly 15,000 public school districts in the country. These Local Education Agencies (LEAs) are responsible for the education of children in their jurisdiction. Each year, they report administrative data, through State Education Agencies (SEAs), to the National Center for Education Statistics (NCES). NCES reviews the data provided and, in some cases, obtains revised data that more accurately reflect the status of LEAs. Information about the roughly 90,000 individual public schools in these local agencies is also collected and maintained by NCES and is available in the CCD Public School Universe Files.

While the data in the basic CCD nonfiscal survey files are a valuable resource for education policy-makers, the utility of the data for policy research has been limited by the presence of missing data and of anomalous values, many of which are clearly erroneous reports. The 13-year Longitudinal CCD Nonfiscal Survey Database is designed to support the research uses of the CCD by enhancing the quality of the data. It is based on the CCD local agency and school files for the school years 1986-87 through 1998-99. These thirteen years saw the end of declining enrollments and a steady increase in enrollments during the 1990s. They also saw the expansion of the Federal State Cooperative Data System and with it the standardization of reporting school district administrative information. Each year there have been increases in overall accuracy and completeness of reporting, so that the strong correlations of measures between years have enabled the implementation of powerful editing and imputation procedures. As a result, the longitudinal files can support valid and reliable studies of school district trends.

Structure of the Longitudinal CCD Nonfiscal Survey Database

The Longitudinal Database consists of three sets of files: (1) a set of 13 local agency-level files, each containing a record for every regular public school district in the United States and its territories; (2) a set of 13 school-level files, each containing a record for every public school in those districts; and (3) a supplementary file of local agency-level information on categories of staffing. Each of the two sets of 13 files consists of thirteen single-year files, one file for each school year, from 1986-87 through 1998-99, plus a single overall file combining quantitative information across the 13 years, the Longitudinal District File and the Longitudinal School File. The supplementary staffing data are contained in a single file with information for the seven school years from 1992-93 through 1998-99.

The single-year files in the Longitudinal Database contain the directory information as recorded in the original CCD nonfiscal survey files maintained by the National Center for Education Statistics (e.g., school name, address, and telephone number). They also contain quantitative fields (e.g., enrollment); and all missing values in these quantitative fields have been filled in (imputed) based on statistical procedures. In addition, clearly erroneous values have been

---

[1] The source files for adding the 1998-99 data to the longitudinal database are the previous longitudinal files, plus "ccdagn98.sd2" (7/20/00) and "sc981c.sd2" (9/12/02). The former differs from "ag981c.sd2" (6/12/01) only in membership counts for Puerto Rico and 13 districts in Massachusetts. In the case of the Massachusetts districts, longitudinal editing yielded virtually the same changes as between the two CCD source files. For Puerto Rico, changes were made to the longitudinal database to match the newer source file.

replaced with values based on the same procedures.[2] Thus, trend data can be graphed and interpreted with greater confidence than before.

The files are linked by a 7-digit code number for each district (LEAID) and a 12-digit code number for each school (MASTERID), the first 7 digits of which are the district code number for the district designated for the school. In the vast majority of cases, the school code number is the same as the original CCD identifier for the school (NCESSCH). However, in cases in which a school changes district designation (e.g., as part of a consolidation), MASTERID remains constant, while the original CCD identifier, NCESSCH, changes.

Cases included in the files

While most of the records in the basic CCD local agency (i.e., district) files refer to entities that we all recognize as school districts, roughly 1,000 of the records refer to different kinds of agencies that are in one manner or another responsible for the education of children. These include agencies that operate in correctional institutions, schools for blind and deaf children, agencies that provide special services to schools in several districts in a region, and administrative agencies that only serve students indirectly. The Longitudinal District File only contains records for "regular" school districts which report employing teachers and enrolling students.[3]

The Longitudinal School File contains records for between 84,000 and 90,000 schools each year, a total of 105,793 schools over the 13-year period. The Longitudinal School File is organized by district, with records for all schools in the same district grouped together. This means that records for schools that changed districts appear more than once on the Longitudinal School File. Thus, the file of 105,793 schools includes 107,317 records, as shown in table 1.

Table 1. Number of schools appearing in different numbers of districts between 1986-87 and 1998-99.

	Number of districts with which the school has been affiliated			
	1	2	3	4
Number of schools	104,282	1,500	9	2
Cumulative number of records	104,282	107,282	107,309	107,317

---

[2] An example of a "clearly erroneous value" would be an enrollment count of 3,000 in 1994-95, in a district that reported enrollments in the two adjacent years of 310 and 320 students and numbers of teachers for the three years of 16, 16, and 18. In that case, a random value in a range around 315 would be imputed for the 1994-95 enrollment.

[3] The district "type code" on CCD takes on values 1 through 7. Regular school districts are normally types 1 and 2, although in some cases, in some states, and in some years, regular school districts are reported to have other type codes. For example, in Minnesota, the result of consolidation of adjacent regular school districts has been labeled as a "regional" district. For 1998-99 24 BIA districts (type 6) were added to the file; and a few districts of type 5 (state institutions) are included because they were classified as regular in a different year. Detailed documentation of CCD can be obtained from the NCES webpage (/nces.ed.gov/ccd/).

All regular districts in the 50 states and the District of Columbia are included in the longitudinal files. In addition, 24 Local Education Agencies (LEAs) are included for the Bureau of Indian Affairs; and one LEA is included for each of the five outlying areas, Puerto Rico, the U.S. Virgin Islands, Guam, the Commonwealth of the Northern Mariana Islands, and American Samoa. The total number of school districts included in the Longitudinal District File, shown by region in table A1, changes from year to year as school districts consolidate, reorganize, and split.

Measures included on the longitudinal files

For every public school district, the basic CCD local agency files contain (1) directory information (name, address, phone number, type, the state's district identification code), (2) a few categorical characteristics, such as grade span and the urbanicity of the community, and (3) counts of schools, students, teachers, special education students, graduates, and dropouts. Other information, such as racial/ethnic distributions and counts of students eligible for the federal free and reduced price lunch program, can be aggregated from the CCD school file to the district level.

The Longitudinal District File includes all the measures on the primary CCD district file except dropout counts, although editing and imputation were only carried out for a subset of the measures. In addition, the Longitudinal District File includes student race/ethnicity data and free and reduced price lunch eligible counts derived from the CCD school file and other sources, as well as two kinds of linkage measure: (1) between years, for districts that consolidate, and (2) between separate elementary and secondary districts. (The linkage between secondary districts and their respective "feeder" elementary districts is provided for only the 1992-93 school year.). The measures that were either edited and imputed or were created for the Longitudinal District File are shown in table 2.

The core measures, SCH (number of schools), MEMBER (total enrollment), and FTE (full-time equivalent teachers), are available for the entire thirteen-year period, from 1986-87 to 1998-99. However, PK12 (prekindergarten-grade 12 enrollment), UG (ungraded enrollment), SPECED (special education students), REGDPL, OTHDPL, OTHHSC, and HSEQVR (recipients of regular and other diplomas, other high school completers, and high school equivalency recipients, respectively), FLE (free-lunch eligible students), ASIAN, WHITE, HISP, BLACK, and AMIND (American Indian and Alaska Native) are not available for the first year, 1986-87. PPOV90 and PPOV95 (percentage of children in poverty) are based solely on 1990 U.S. Census data, with 1995 updates; HSEQVR is not available after 1990-91; and SECLEA (secondary district for which an elementary district is a "feeder") is created only for 1992-93. Starting in 1998-99, REGDPL and OTHDPL are combined into a single set of counts, for each gender and race/ethnicity.

Although a single primary data source is indicated for each measure, other sources were used both to identify apparent wrong responses and to generate statistical imputations for missing or apparently wrong responses. For example, information on expenditures for school lunch from the F33 School District Fiscal Survey was used in the imputation of missing percentages of free lunch eligible students.

4

Table 2. Measures examined or created for the CCD Longitudinal District File

Variable Name	Definition	Primary Source
SCH	Number of schools	CCD Agency File
MEMBER	Total enrollment	CCD Agency File
FTE	Total number of full-time equivalent teachers	CCD Agency File
PK12	Number of students in grades PK through 12	CCD Agency File
UG	Number of ungraded students	CCD Agency File
SPECED	Number of special education students	CCD Agency File
REGDPL	Number of regular diplomas awarded in past year	CCD Agency File
OTHDPL	Number of other diplomas awarded in past year	CCD Agency File
OTHHSC	No. of other high school completers in past year	CCD Agency File
HSEQVR	Number of high school equivalency recipients in past year	CCD Agency File
GSLO	Lowest grade with pupils enrolled	CCD Agency File
GSHI	Highest grade with pupils enrolled	CCD Agency File
LOCALE	Community type of most schools in district	CCD School Aggregate
FLE	No. of students eligible for free lunch program	CCD School Aggregate
ASIAN, WHITE, BLACK, HISP, AMIND	Counts of race/ethnic groups in enrollment	CCD School Aggregate
PPOV90,95	Fraction of school aged children in poverty	1990 U.S.Census
NXTYRID	For closing districts, NCES ID of district receiving most of its students next year.	New
PRVYRID	For districts receiving students from a closed district, the NCES ID of the closed district.	New
SECLEA	For elementary districts, the NCES ID of the district receiving most of its students for secondary education	New
YRS	String variable: i-th character, for year (1985/86+i), is Y (LEA open, w schools), N (LEA but no schools), or M (LEA not open)	New

New variables added for 1998-99

Variable Name	Definition	Primary Source
REDLCH	No. of students eligible for reduced price lunch program	CCD School Aggregate
TOTFRL	FLE + REDLCH	CCD School Aggregate
LEP	No. of students with limited English proficiency	CCD Agency File
MALDPL, FEMDPL	Male, female diploma recipients	CCD Agency File
AMDPL, ASDPL, BLDPL, HIDPL, WHDPL	Diploma recipients, by race/ethnicity	CCD Agency File
MALOHC, FEMOHC	Male, female other high school completers	CCD Agency File
AMOHC, ASOHC, BLOHC, HIOHC, WHOHC	Other high school completers, by race/ethnicity	CCD Agency File

Note: PK12, UG, SPECED, REGDPL, OTHHSC, ASIAN, WHITE, HISP, BLACK, AMIND, FLE are not available for 1986-87. PPOV based solely on 1990 and 1995 data. HSEQVR is available only for 1987-88 through 1990-91. SECLEA is available only for 1992-93.

The Longitudinal School File contains information on the number of students enrolled, MEMBER86 through MEMBER98, on the number of full-time equivalent teachers, FTE86 through FTE98, on race/ethnicity counts, White, Black, Hispanic, Asian, and American Indian/Alaska Native, and on free and reduced price lunch eligibility counts for each school in the districts on the Longitudinal District File.[4] The file also includes a unique MASTER ID for each school, as shown in table 3. A school appearing in different districts in different years is represented by multiple records on the Longitudinal School File, with the same MASTER ID but a different LEA ID. Other measures currently on the CCD Public School Universe File and to be added to the Longitudinal School File in the future include grade-by-grade enrollments.

The third component in the Longitudinal CCD Non-Fiscal Survey Database is the Longitudinal District Staff File. For the years 1992-93 through 1998-99, it contains longitudinally edited counts of prekindergarten, kindergarten, elementary, secondary, and ungraded teachers, teacher aides, instructional coordinators, elementary and secondary guidance counselors, library specialists and support staff, school and district administrators and support staff, and student and other support staff.

Table 3. Measures examined or created for the CCD Longitudinal School File

Variable Name	Definition	Primary Source
MASTER ID	Permanent school identifier	New
LEAID	District identifier	CCD Agency File
NCESSCH	Current school identifier	CCD School File
MEMBER	Number of enrolled students	CCD School File
FTE	Number of full-time equivalent teachers	CCD School File
ASIAN	Number of Asian students	CCD School File
BLACK	Number of Black students	CCD School File
HISP	Number of HISP students	CCD School File
IND	Number of American Indian students	CCD School File
WHITE	Number of White students	CCD School File
FLE	No. of students eligible for free lunch program	CCD School Aggregate
REDLCH	No. of students eligible for reduced price lunch program	CCD School Aggregate
TOTFRL	FLE + REDLCH	CCD School Aggregate
LYRS	String variable: i-th character, for year (1985/86+i), is Y (School open in this LEA), or N (School open in another LEA or closed)	New
SYRS	String variable: i-th character, for year (1985/86+i), is Y (School open), or N (School closed)	New

---

[4] Prior to 1998-99, free lunch eligible counts are provided on CCD. Starting in 1998-99, total free and reduced price lunch eligibility counts are broken down in free lunch counts and reduced price lunch counts.

The editing and imputation procedures for adding the 1998-99 CCD data to the Longitudinal Database are adapted from the procedures used in earlier rounds of the development of the database. The procedures are described in Appendix B, which is divided into three sections, corresponding to the three previous rounds of editing and imputation (labeled rounds 2, 3, and 4).

The greatest difficulty in adding the 1998-99 data centered on the addition of new charter schools/districts (about 180 of the 279 new districts are charter schools). The variety of organizational systems for charter schools in different states creates confusion in the linkage of schools to districts. In some states, CCD records each charter school as a separate district; while in other states charter schools are counted as new schools in existing districts. In at least one state, Arizona, charter schools frequently change their association with sponsoring districts, and these sponsoring districts may be geographically distant from the charter schools.

The frequencies of imputed values on the Longitudinal Database for 1998-99 are shown in Table 4, for districts, and Table 5, for schools.

Table 4. Frequency of districts with imputed values for 1998-99 on the Longitudinal District File

Variable Name	Reported Values	Imputed Values
SCH	14,809	15
MEMBER	14,742	82
FTE	14,443	381
PK12	14,571	253
UG	10,385	4,439
SPECED	14,656	168
GSLO	14,705	119
GSHI	14,718	106
LOCALE	14,814	10
ASIAN, BLACK, HISP, AMIND, WHITE	14,173	651
FLE	11,801	3,023
REDLCH	8,034	6,790
TOTFRL	8,034	6,790
LEP	8,268	6,556
TOTDPL	13,573	1,251
OTHHSC	2,097	12,727
MALDPL, FEMDPL	7,981	6,843
AMDPL, ASDPL, BLDPL, HIDPL, WHDPL	8,311	6,513
MALOHC, FEMOHC	1,151	13,673
AMOHC, ASOHC, BLOHC, HIOHC, WHOHC	1,159	13,665

Table 5. Frequency of schools with imputed values for 1998-99 on the Longitudinal School File

Variable Name	Reported Values	Imputed Values
MEMBER	89,825	1,467
FTE	87,099	4,193
ASIAN, BLACK, HISPANIC, IND, WHITE	88,226	3,066
FLE, RDLCH, TOTFRL	56,558	34,734

The results of the editing and imputation for 1998-99 are included in the tables in Appendix A. For nearly all measures on CCD, the changes from 1997-98 to 1998-99 follow smooth trends. The major exceptions to this are the following.

- Ungraded students: In 1998-99, Illinois essentially stopped classifying students as ungraded. In the same year, Kentucky began to classify primary grade students as ungraded, but these are counted in CCD as being enrolled in grades 1, 2, and 3.

- Other high school completers: In 1998-99, New York, Michigan, Iowa, and Puerto Rico dramatically reduced their counts of other high school completers. These counts were changed by a factor of more than two in several other states, as well.

Appendix A: Codebook

The information in the Longitudinal CCD Non-Fiscal Survey Database, which consists of the fields listed in tables 2 and 3, is stored in two forms: (1) on a separate set of files, one for each year, merged with unedited CCD directory information, and (2) on a single file, containing information for all years, but without directory information. The files can be merged using the common identifiers, LEAID (for districts) and MASTERID (for schools). Variables have the same names on both the individual year files and the combined file. The stems of the variable names are as given in tables 2 and 3, to which a two-digit representation of the year is added. Thus, for example, MEMBER95 is the membership count for a district or a school in the 1995-96 school-year.

Frequency distributions of these fields, based on the Longitudinal District File, are given in tables A1 through A11, and based on the Longitudinal School File in tables A12 through A15. Tables A16 through A18 are based on the Longitudinal District Staffing File. For tables A1 through A9 and A12 through A14, the frequencies are broken down by region of the country, defined by:

Northeast: Maine, New Hampshire, Vermont, New York, Massachusetts, Rhode Island, Connecticut, New Jersey, Pennsylvania

South: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Alabama, Mississippi, Louisiana, Arkansas, Oklahoma, Texas

Midwest: Michigan, Wisconsin, Minnesota, North Dakota, South Dakota, Ohio, Indiana, Illinois, Iowa, Nebraska, Missouri, Kansas

West: Montana, Idaho, Washington, Alaska, Wyoming, Colorado, Utah, New Mexico, Nevada, Arizona, Oregon, California, Hawaii

Territories: American Samoa, Guam, Marianas, Puerto Rico, Virgin Islands, and starting in 1998-99, Bureau of Indian Affairs

Table A1. Number of regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87	2,981	3,406	5,981	2,987	5	15,360
1987-88	2,978	3,391	5,930	2,985	5	15,289
1988-89	2,967	3,387	5,881	2,975	5	15,215
1989-90	2,964	3,373	5,833	2,960	5	15,135
1990-91	2,970	3,352	5,780	2,948	5	15,055
1991-92	2,965	3,324	5,720	2,939	5	14,953
1992-93	2,966	3,304	5,630	2,911	5	14,816
1993-94	2,959	3,273	5,544	2,868	5	14,649
1994-95	2,939	3,263	5,491	2,827	5	14,525
1995-96	2,960	3,259	5,519	2,806	5	14,549
1996-97	2,956	3,274	5,554	2,864	5	14,653
1997-98	2,974	3,287	5,548	2,849	5	14,663
1998-99	3,009	3,359	5,552	2,875	29	14,824

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A2. Number of schools in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87	13,809	27,181	24,912	16,482	1,865	84,249
1987-88	13,847	26,988	24,879	16,544	1,883	84,141
1988-89	13,876	27,063	24,530	16,767	1,803	84,039
1989-90	13,886	27,199	24,607	16,920	1,788	84,400
1990-91	13,983	27,289	24,602	17,722	1,743	85,339
1991-92	13,855	27,205	24,615	17,930	1,709	85,314
1992-93	13,890	27,390	24,600	17,792	1,714	85,386
1993-94	13,939	27,561	24,927	17,939	1,707	86,073
1994-95	14,040	27,901	24,992	18,157	1,689	86,779
1995-96	14,066	28,240	25,315	18,328	1,685	87,634
1996-97	14,147	28,606	25,501	18,774	1,682	88,710
1997-98	14,259	29,040	25,646	19,159	1,700	89,804
1998-99	14,360	29,848	25,708	19,520	1,856	91,292

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A3. Number of students in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87	7,175,364	14,295,003	9,846,574	8,290,007	741,666	40,348,614
1987-88	7,135,280	14,349,705	9,860,720	8,479,351	738,655	40,563,711
1988-89	7,095,194	14,485,672	9,728,807	8,674,644	727,770	40,712,087
1989-90	7,091,424	14,517,722	9,832,625	8,895,755	715,198	41,052,724
1990-91	7,176,224	14,798,071	9,897,516	9,144,325	710,025	41,726,161
1991-92	7,302,835	15,031,613	10,013,102	9,473,952	712,542	42,534,044
1992-93	7,430,361	15,310,012	10,154,159	9,684,360	712,078	43,290,970
1993-94	7,547,826	15,542,238	10,252,239	9,867,845	707,507	43,917,655
1994-95	7,662,120	15,789,760	10,339,065	10,042,856	698,449	44,532,250
1995-96	7,794,802	16,052,829	10,459,059	10,257,713	698,376	45,262,779
1996-97	7,895,603	16,320,036	10,600,980	10,516,862	697,449	46,030,930
1997-98	7,974,170	16,517,651	10,638,124	10,693,348	695,482	46,518,775
1998-99	8,037,379	16,667,151	10,671,353	10,863,298	740,524	46,979,705

Note: Source: U.S. Dept. of Education, Natl. Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A4. Number of full-time equivalent teachers in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87	447,122	793,504	559,715	386,689	35,881	2,222,912
1987-88	455,403	818,357	561,792	396,901	36,923	2,269,376
1988-89	458,913	839,430	561,439	402,838	37,013	2,299,633
1989-90	461,488	859,387	583,333	416,339	37,447	2,357,993
1990-91	469,390	876,591	577,786	433,017	38,401	2,395,186
1991-92	467,810	891,171	580,710	437,715	41,381	2,418,786
1992-93	477,589	888,304	590,214	451,303	42,754	2,450,164
1993-94	485,121	932,939	592,807	456,805	44,117	2,511,789
1994-95	493,030	949,397	602,816	465,867	44,348	2,555,458
1995-96	499,346	973,167	610,684	475,603	44,506	2,603,306
1996-97	506,240	990,867	624,504	499,972	44,050	2,665,632
1997-98	521,575	1,014,956	633,996	526,624	43,143	2,740,294
1998-99	538,300	1,052,877	645,427	544,659	48,900	2,830,163

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A5. Number of special education/IEP students in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	721,550	1,524,915	986,133	728,750	16,445	3,977,793
1988-89	742,925	1,602,270	979,949	768,322	18,355	4,111,821
1989-90	792,407	1,596,812	1,081,230	788,399	19,592	4,278,440
1990-91	776,489	1,679,681	1,059,658	858,714	18,718	4,393,260
1991-92	791,585	1,665,628	1,133,582	893,949	18,384	4,503,128
1992-93	858,415	1,838,088	1,174,507	919,053	18,110	4,808,173
1993-94	833,572	1,902,307	1,201,498	983,642	18,122	4,939,141
1994-95	846,135	1,957,634	1,050,605	990,447	17,971	4,862,792
1995-96	876,760	1,971,523	1,008,075	1,046,826	17,823	4,921,007
1996-97	939,926	2,008,052	1,034,157	1,103,580	46,104	5,131,819
1997-98	977,408	2,085,561	1,220,304	1,146,253	55,244	5,484,770
1998-99	1,008,307	2,161,102	1,245,836	1,203,750	66,418	5,685,413

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A6. Number of free lunch eligible students in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	1,599,885	4,473,860	1,985,030	2,182,996	580,664	10,822,435
1988-89	1,606,585	4,682,761	1,945,387	2,325,340	598,788	11,158,861
1989-90	1,633,931	4,751,232	1,992,924	2,495,893	550,721	11,424,701
1990-91	1,701,659	4,933,876	2,087,168	2,676,597	573,329	11,972,629
1991-92	1,788,974	4,924,164	2,225,753	2,901,525	648,361	12,488,777
1992-93	1,865,246	5,419,868	2,312,895	3,118,200	523,439	13,239,648
1993-94	1,948,001	5,498,947	2,323,330	3,300,706	509,955	13,580,939
1994-95	1,999,437	5,698,185	2,338,034	3,452,196	540,875	14,028,727
1995-96	2,047,743	5,867,991	2,449,755	3,646,173	534,870	14,546,532
1996-97	2,230,903	6,083,140	2,509,510	3,791,876	547,962	15,163,391
1997-98	2,267,055	6,161,711	2,568,451	3,928,398	540,331	15,465,946
1998-99	2,222,829	5,930,573	2,456,229	3,489,348	552,483	14,651,462

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A7. Number of ungraded students in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	242,123	121,047	382,548	123,395	497	869,610
1988-89	236,733	138,087	256,510	158,047	13,861	803,238
1989-90	236,827	149,794	357,579	146,903	13,123	904,226
1990-91	235,742	147,814	311,554	151,216	12,377	858,703
1991-92	235,500	121,336	402,123	158,112	12,950	930,021
1992-93	233,360	159,110	326,469	148,815	13,685	881,439
1993-94	238,979	159,267	353,638	130,663	14,892	897,439
1994-95	238,185	135,717	353,465	126,342	1,945	855,654
1995-96	238,434	136,981	353,059	126,867	15,525	870,866
1996-97	242,397	110,635	383,957	131,694	14,670	883,353
1997-98	252,434	114,406	394,568	113,997	13,914	889,319
1998-99	259,660	106,835*	132,308	114,115	15,196	628,114

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.
* In addition, Kentucky reported 148,142 K-3 students as ungraded in 1998-99.

Table A8a. Number of regular diplomas earned in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	464,145	814,777	605,329	418,082	32,958	2,335,291
1988-89	491,139	828,662	690,374	466,910	34,566	2,511,651
1989-90	465,838	827,761	667,172	455,823	34,379	2,450,973
1990-91	436,897	784,691	628,884	449,764	32,277	2,332,513
1991-92	408,239	769,719	584,592	450,352	32,068	2,244,970
1992-93	408,675	741,473	576,003	452,413	32,274	2,210,838
1993-94	407,109	754,613	583,145	470,919	31,860	2,247,646
1994-95	401,122	744,025	568,117	481,503	31,810	2,226,577
1995-96	405,373	770,078	585,135	493,810	30,714	2,285,110
1996-97	406,271	767,537	588,682	497,487	32,108	2,292,086
1997-98	419,361	782,438	606,050	520,402	32,697	2,360,948

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

13

Table A8b. Number of other diplomas earned in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	549	10,692	6,437	8,205	5,568	31,451
1988-89	1,599	12,972	11,721	902	4,559	31,753
1989-90	622	12,871	10,353	795	4,119	28,760
1990-91	1,025	23,925	9,668	5,202	4,970	44,790
1991-92	1,440	16,861	4,578	2,471	11,032	36,382
1992-93	4,626	32,266	4,296	5,708	24,640	71,536
1993-94	4,608	16,537	5,135	2,829	10,312	39,421
1994-95	6,075	14,057	5,247	2,759	9,728	37,866
1995-96	6,231	11,501	4,878	2,891	15,544	41,045
1996-97	5,941	9,906	3,688	3,217	14,712	37,464
1997-98	5,763	10,997	3,977	4,451	14,826	40,014

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A8c. Number of other high school completers in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	955	5,041	808	1,587	53	8,444
1988-89	2,344	10,025	993	1,899	2,471	17,732
1989-90	2,602	8,958	921	1,819	2	14,302
1990-91	2,575	11,871	1,480	2,936	7	18,869
1991-92	1,518	10,043	814	2,656	41	15,072
1992-93	2,661	11,806	2,358	4,247	2,343	23,415
1993-94	2,880	13,341	2,616	4,923	6,929	30,689
1994-95	3,102	14,271	3,175	5,271	2,268	28,087
1995-96	2,676	17,493	3,352	5,957	6,956	36,434
1996-97	2,896	20,418	3,869	6,924	9,159	43,266
1997-98	2,722	23,604	3,418	6,427	11,898	48,069
1998-99	173	26,085	862	5,277	11	32,408

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A9a. Number of Asian students in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	196,190	185,002	133,900	657,379	39,149	1,211,620
1988-89	204,369	191,422	138,028	686,851	40,081	1,260,751
1989-90	216,438	201,105	146,993	718,693	39,568	1,322,797
1990-91	229,383	214,979	155,252	745,213	41,368	1,386,195
1991-92	243,520	226,509	161,821	787,615	44,609	1,464,074
1992-93	257,612	241,957	170,508	817,723	48,132	1,535,932
1993-94	270,318	254,876	179,634	842,536	49,779	1,597,143
1994-95	279,487	266,794	187,971	862,757	50,888	1,647,897
1995-96	293,928	280,493	195,964	887,803	52,184	1,710,372
1996-97	308,584	294,587	205,032	911,126	54,699	1,774,028
1997-98	319,497	305,631	213,578	933,800	55,562	1,828,068
1998-99	332,701	319,931	221,542	952,047	56,187	1,882,408

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A9b. Number of Black students in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	1,047,681	3,715,780	1,322,683	540,327	20,104	6,646,575
1988-89	1,046,878	3,745,597	1,289,705	546,278	20,074	6,648,532
1989-90	1,049,434	3,749,103	1,308,440	553,247	17,846	6,678,070
1990-91	1,066,112	3,814,252	1,314,317	566,763	18,489	6,779,933
1991-92	1,089,757	3,863,210	1,323,037	588,175	19,203	6,883,382
1992-93	1,117,156	3,956,757	1,344,846	604,925	20,131	7,043,815
1993-94	1,142,360	4,036,961	1,367,214	619,645	19,946	7,186,126
1994-95	1,163,979	4,131,235	1,387,929	637,611	19,983	7,340,737
1995-96	1,189,612	4,218,772	1,422,294	656,943	19,327	7,506,948
1996-97	1,206,601	4,303,327	1,469,250	676,438	18,476	7,674,092
1997-98	1,231,602	4,364,766	1,493,460	695,787	18,894	7,804,509
1998-99	1,241,238	4,419,112	1,525,482	708,414	17,904	7,912,150

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A9c. Number of Hispanic students in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	625,539	1,268,679	281,030	1,789,564	676,096	4,640,908
1988-89	640,714	1,331,384	282,776	1,908,662	664,411	4,827,947
1989-90	659,629	1,368,242	307,238	2,064,287	654,333	5,053,729
1990-91	693,099	1,475,529	323,497	2,221,876	647,300	5,361,301
1991-92	726,161	1,550,212	343,430	2,363,941	645,296	5,629,040
1992-93	761,387	1,622,925	363,685	2,477,695	640,265	5,865,957
1993-94	796,519	1,703,707	383,441	2,589,601	634,633	6,107,901
1994-95	829,117	1,788,678	405,304	2,701,002	624,487	6,348,588
1995-96	870,547	1,887,456	434,097	2,838,226	624,344	6,654,670
1996-97	899,724	1,997,170	466,534	3,001,660	622,162	6,987,250
1997-98	924,179	2,083,065	491,771	3,126,235	619,689	7,244,939
1998-99	951,064	2,184,770	517,316	3,263,819	616,857	7,533,826

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A9d. Number of American Indian and Alaska Native students in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	13,956	103,937	63,829	181,861	13	363,596
1988-89	12,431	105,880	64,039	187,978	10	370,338
1989-90	13,015	108,065	68,560	193,848	7	383,495
1990-91	13,866	114,035	69,621	198,674	21	396,217
1991-92	16,273	120,169	72,546	213,400	22	422,410
1992-93	17,194	126,427	75,397	219,429	40	438,487
1993-94	17,736	132,522	79,478	228,955	30	458,721
1994-95	19,275	139,061	81,732	236,716	93	476,877
1995-96	20,893	146,431	84,335	244,275	33	495,967
1996-97	22,835	151,756	87,185	254,236	37	516,049
1997-98	23,937	154,478	89,130	257,351	35	524,931
1998-99	21,556	161,811	89,517	260,179	48,561	581,624

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A9e. Number of White students in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	5,251,914	9,076,307	8,059,278	5,310,220	3,293	27,701,012
1988-89	5,190,802	9,111,389	7,954,259	5,344,875	3,194	27,604,519
1989-90	5,152,908	9,091,207	8,001,394	5,365,680	3,444	27,614,633
1990-91	5,173,764	9,179,276	8,034,829	5,411,799	2,847	27,802,515
1991-92	5,227,124	9,271,513	8,112,268	5,520,821	3,412	28,135,138
1992-93	5,277,012	9,361,946	8,199,723	5,564,588	3,510	28,406,779
1993-94	5,320,893	9,414,172	8,242,472	5,587,108	3,119	28,567,764
1994-95	5,370,262	9,463,992	8,276,129	5,604,770	2,998	28,718,151
1995-96	5,419,822	9,519,677	8,322,369	5,630,466	2,488	28,894,822
1996-97	5,457,859	9,573,196	8,372,979	5,673,402	2,075	29,079,511
1997-98	5,474,955	9,609,711	8,350,185	5,680,175	1,302	29,116,328
1998-99	5,490,820	9,581,527	8,317,496	5,678,839	1,015	29,069,697

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A10. Number of regular public school districts, by locale type and year

	Large Central City	Midsize Central City	Large City Fringe	Midsize City Fringe	Large Town	Small Town	Rural outside MSA	Rural inside MSA
1986-87	171	762	1,247	879	240	4,613	7,447	-
1987-88	162	756	1,245	870	236	4,594	7,425	-
1988-89	162	752	1,243	870	225	4,560	7,402	-
1989-90	160	725	1,307	913	222	4,388	7,419	-
1990-91	160	734	1,319	943	230	4,344	7,324	-
1991-92	176	721	1,292	910	343	4,318	7,192	-
1992-93	170	719	1,290	911	272	4,291	7,162	-
1993-94	173	710	1,286	909	328	4,290	6,952	-
1994-95	240	843	2,369	960	188	2,607	7,317	-
1995-96	249	818	2,769	1,179	168	2,375	6,990	-
1996-97	307	851	2,778	1,186	176	2,399	6,955	-
1997-98	326	877	2,777	1,189	180	2,390	6,923	-
1998-99	358	704	2,999	1,617	118	2,286	5,623	1,119

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.
Note: Rural districts within Metropolitan Statistical Areas were not distinguished from other districts before 1998-99.

Table A11a. Number of regular public school districts, by lowest grade and year

	PK	KG	1	2	3	4	5	6	7	8	9	10	11	12
1986-87	2,710	11691	278	4	1	5	10	4	122	2	515	4	1	0
1987-88	2,901	11230	411	36	15	10	15	7	123	5	511	7	1	0
1988-89	2,988	11122	352	42	17	13	12	14	120	5	508	7	1	0
1989-90	3,451	10640	292	38	15	19	9	12	124	6	510	6	1	0
1990-91	4,014	10050	250	41	16	14	13	11	117	5	507	1	1	0
1991-92	4,899	9,066	243	37	18	9	15	14	119	4	508	1	1	0
1992-93	5,093	8,775	194	42	20	11	14	12	121	5	495	3	1	0
1993-94	5,267	8,484	178	41	14	14	7	13	117	3	464	3	1	0
1994-95	5,475	8,185	172	30	20	13	9	12	116	5	442	3	1	0
1995-96	5,525	8,136	174	43	13	10	10	20	116	5	446	3	4	2
1996-97	5,886	7,818	165	35	16	13	14	33	129	13	466	6	6	2
1997-98	5,961	7,774	147	21	18	9	15	43	134	14	459	4	8	4
1998-99	6,734	7,220	77	6	6	7	20	65	143	9	472	5	7	2

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A11b. Number of regular public school districts, by highest grade and year

	PK	KG	1	2	3	4	5	6	7	8	9	10	11	12
1986-87	1	1	2	1	8	18	14	552	29	3,108	29	2	6	11574
1987-88	2	3	7	8	24	49	85	728	149	2,711	39	3	3	11461
1988-89	2	1	6	10	16	41	84	745	138	2,693	34	2	5	11424
1989-90	2	3	3	5	16	40	90	732	146	2,668	37	5	6	11370
1990-91	2	3	4	5	13	41	80	736	126	2,670	37	5	9	11309
1991-92	2	1	2	5	12	37	81	739	127	2,622	32	2	12	11260
1992-93	2	1	3	7	14	32	74	715	128	2,552	27	6	13	11212
1993-94	1	1	5	3	12	41	65	696	94	2,473	27	4	9	11175
1994-95	1	1	1	5	4	33	85	655	112	2,389	27	4	11	11154
1995-96	1	2	2	8	18	42	86	659	105	2,364	24	11	16	11169
1996-97	1	3	4	8	21	46	109	636	101	2,368	41	13	29	11222
1997-98	1	2	5	9	23	42	114	644	95	2,342	27	25	21	11261
1998-99	1	4	2	7	21	27	81	593	52	2,532	38	27	37	11352

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A12.   Number of schools in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87	13,809	27,181	24,912	16,482	1,865	84,249
1987-88	13,847	26,988	24,879	16,544	1,883	84,141
1988-89	13,876	27,063	24,530	16,767	1,803	84,039
1989-90	13,886	27,199	24,607	16,920	1,788	84,400
1990-91	13,983	27,289	24,602	17,722	1,743	85,339
1991-92	13,855	27,205	24,615	17,930	1,709	85,314
1992-93	13,890	27,390	24,600	17,792	1,714	85,386
1993-94	13,939	27,561	24,927	17,939	1,707	86,073
1994-95	14,040	27,901	24,992	18,157	1,689	86,779
1995-96	14,066	28,240	25,315	18,328	1,685	87,634
1996-97	14,147	28,606	25,501	18,774	1,682	88,710
1997-98	14,259	29,040	25,646	19,159	1,700	89,804
1998-99	14,360	29,848	25,708	19,520	1,856	91,292

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal School File.

Table A13.   Number of students in schools in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87	7,186,039	14,286,915	9,830,699	8,280,585	741,666	40,325,904
1987-88	7,116,737	14,371,497	9,829,635	8,472,860	738,087	40,528,816
1988-89	7,094,093	14,484,006	9,788,353	8,666,601	730,286	40,763,339
1989-90	7,089,565	14,571,031	9,793,701	8,887,947	715,348	41,057,592
1990-91	7,175,215	14,748,831	9,868,363	9,185,465	708,989	41,686,863
1991-92	7,299,950	15,063,771	10,000,734	9,468,942	713,620	42,547,017
1992-93	7,428,656	15,303,570	10,135,155	9,687,995	712,753	43,268,129
1993-94	7,546,366	15,542,957	10,238,690	9,866,065	707,803	43,901,881
1994-95	7,660,801	15,812,991	10,335,207	10,031,936	698,449	44,539,384
1995-96	7,796,778	16,049,409	10,468,927	10,244,464	706,050	45,265,628
1996-97	7,902,145	16,300,679	10,567,881	10,509,834	699,064	45,979,603
1997-98	7,979,621	16,499,324	10,641,456	10,693,424	695,986	46,509,811
1998-99	8,020,768	16,649,495	10,655,545	10,864,580	740,533	46,930,921

Note: Source: U.S. Dept. of Education, Natl. Center for Education Statistics, Common Core of Data, Longitudinal School File.

Table A14. Number of full-time equivalent teachers in schools in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87	446,236	791,613	551,448	385,073	35,880	2,210,250
1987-88	455,201	817,070	554,363	396,382	36,923	2,259,938
1988-89	458,929	848,744	553,724	402,387	37,373	2,301,157
1989-90	461,476	859,435	583,271	416,249	37,447	2,357,878
1990-91	468,746	875,600	570,515	431,825	38,401	2,385,086
1991-92	467,651	890,987	574,551	437,459	41,381	2,412,029
1992-93	461,837	895,092	585,748	443,414	42,985	2,429,076
1993-94	470,030	933,606	584,132	450,433	43,859	2,482,060
1994-95	477,418	953,698	594,839	460,100	44,129	2,530,185
1995-96	483,368	971,342	604,431	470,655	43,883	2,573,678
1996-97	494,605	985,309	616,639	495,190	43,951	2,635,694
1997-98	507,181	1,009,779	627,337	521,498	43,127	2,708,921
1998-99	520,600	1,033,858	647,648	540,181	48,739	2,791,026

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal School File. Figures are reported in tenths; therefore, totals may not match sums across regions, due to rounding.

Table A15a. Number of Asian students in schools in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	195,780	185,423	133,093	657,129	39,349	1,210,774
1988-89	204,327	191,333	138,864	686,877	39,996	1,261,397
1989-90	216,424	202,123	145,869	718,305	40,952	1,323,673
1990-91	229,363	214,303	154,157	755,496	41,437	1,394,756
1991-92	243,340	226,716	161,328	787,481	44,347	1,463,212
1992-93	257,481	241,816	170,164	817,729	48,675	1,535,865
1993-94	270,180	254,860	179,474	842,429	50,022	1,596,965
1994-95	279,352	267,118	187,896	862,380	50,888	1,647,634
1995-96	293,707	280,373	196,108	887,492	53,286	1,710,966
1996-97	308,690	294,363	203,783	911,810	54,667	1,773,313
1997-98	319,545	305,574	213,682	933,801	55,462	1,828,064
1998-99	331,732	319,864	221,506	952,152	56,196	1,881,450

Note: School data not available for four territories before 1991. Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal School File.

Table A15b. Number of Black students in schools in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	1,041,229	3,713,158	1,320,241	539,840	20,813	6,635,281
1988-89	1,046,749	3,740,942	1,311,329	545,671	20,625	6,665,316
1989-90	1,049,279	3,759,526	1,305,739	552,801	18,323	6,685,668
1990-91	1,066,023	3,807,387	1,310,362	568,366	19,143	6,771,281
1991-92	1,088,925	3,876,476	1,322,445	588,018	19,709	6,895,573
1992-93	1,116,580	3,955,131	1,343,444	605,118	20,155	7,040,428
1993-94	1,141,787	4,037,975	1,362,400	619,598	19,989	7,181,749
1994-95	1,163,416	4,135,888	1,384,462	637,048	19,983	7,340,797
1995-96	1,189,395	4,218,532	1,428,783	656,336	19,328	7,512,374
1996-97	1,207,594	4,295,453	1,466,663	675,836	19,204	7,664,750
1997-98	1,231,615	4,363,080	1,499,959	695,781	18,894	7,809,329
1998-99	1,239,322	4,416,902	1,527,262	708,518	17,904	7,909,908

Note: School data not available for four territories before 1991. Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal School File.

Table A15c. Number of Hispanic students in schools in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	624,073	1,268,662	280,212	1,788,433	674,560	4,635,940
1988-89	640,593	1,330,531	289,294	1,907,617	666,455	4,834,490
1989-90	659,523	1,397,792	306,345	2,063,536	652,611	5,079,807
1990-91	693,049	1,460,791	322,659	2,223,844	645,514	5,345,857
1991-92	725,473	1,549,046	343,382	2,363,593	646,133	5,627,627
1992-93	760,834	1,622,173	363,566	2,479,546	640,270	5,866,389
1993-94	795,932	1,705,035	383,062	2,589,127	634,640	6,107,796
1994-95	828,555	1,791,687	405,095	2,698,843	624,487	6,348,667
1995-96	869,990	1,887,090	434,379	2,836,138	630,846	6,658,443
1996-97	899,402	1,996,877	465,753	3,000,007	623,074	6,985,113
1997-98	924,071	2,082,971	491,923	3,126,238	620,296	7,245,499
1998-99	949,131	2,184,549	517,085	3,264,283	616,857	7,531,905

Note: School data not available for four territories before 1991. Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal School File.

Table A15d. Number of American Indian and Alaska Native students in schools in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	13,937	103,879	63,245	181,813	13	362,887
1988-89	12,432	105,834	63,937	187,731	10	369,944
1989-90	13,008	108,084	68,265	193,655	7	383,019
1990-91	13,870	114,008	69,377	199,864	21	397,140
1991-92	16,265	120,272	72,412	213,010	22	421,981
1992-93	17,190	126,406	75,141	220,088	41	438,866
1993-94	17,730	132,506	79,449	228,668	30	458,383
1994-95	19,274	139,161	81,782	236,271	93	476,581
1995-96	20,888	145,756	84,427	243,262	34	494,367
1996-97	22,823	151,464	86,628	253,608	37	514,560
1997-98	23,930	154,339	89,098	257,348	35	524,750
1998-99	21,514	161,855	89,440	260,144	48,561	581,514

Note: School data not available for four territories before 1991. Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal School File.

Table A15e. Number of White students in schools in regular public school districts, by region and year

	Northeast	South	Midwest	West	Territories	Total
1986-87						
1987-88	5,241,718	9,100,375	8,032,844	5,305,645	3,352	27,683,934
1988-89	5,189,992	9,115,366	7,984,929	5,338,705	3,200	27,632,192
1989-90	5,151,331	9,103,506	7,967,483	5,359,650	3,455	27,585,425
1990-91	5,172,910	9,152,342	8,011,808	5,437,895	2,874	27,777,829
1991-92	5,225,947	9,291,261	8,101,167	5,516,840	3,409	28,138,624
1992-93	5,276,571	9,358,044	8,182,840	5,565,514	3,612	28,386,581
1993-94	5,320,737	9,412,581	8,234,305	5,586,243	3,122	28,556,988
1994-95	5,370,204	9,479,137	8,275,972	5,597,394	2,998	28,725,705
1995-96	5,422,798	9,517,658	8,325,230	5,621,236	2,556	28,889,478
1996-97	5,463,636	9,562,522	8,345,054	5,668,573	2,082	29,041,867
1997-98	5,480,460	9,593,360	8,346,794	5,680,256	1,299	29,102,169
1998-99	5,479,069	9,566,325	8,300,252	5,679,483	1,015	29,026,144

Note: School data not available for four territories before 1991. Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal School File.

Table A16. Number of teachers, by teacher grade level and year

	Prekindergarten	Kindergarten	Elementary	Secondary	Ungraded	Total Teachers
1992-93	14,043	112,870	1,246,940	882,353	194,308	2,450,513
1993-94	15,906	112,638	1,281,338	897,743	204,032	2,511,657
1994-95	16,550	112,839	1,306,015	914,228	210,705	2,560,337
1995-96	18,459	116,026	1,310,436	947,795	210,368	2,603,084
1996-97	19,564	120,405	1,351,125	959,823	216,873	2,667,790
1997-98	20,355	124,023	1,389,895	982,112	230,647	2,747,032
1998-99	21,613	129,160	1,419,720	1,021,928	234,346	2,826,767

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A17. Number of administrators and support staff, by staff type and year

	District Administrator	District Support Staff	School Administrator	School Support Staff	Instructional Coordinators & Supervisors	Other Support Staff
1992-93	45,112	134,328	122,742	182,535	28,584	1,011,451
1993-94	46,369	134,243	122,473	185,259	27,140	1,021,565
1994-95	48,144	126,170	123,489	183,273	29,617	1,007,114
1995-96	47,798	136,246	125,478	173,339	30,281	1,032,007
1996-97	45,119	139,805	127,973	177,519	29,507	1,039,227
1997-98	48,412	138,856	130,294	182,178	31,879	1,087,037
1998-99	49,270	138,874	134,235	190,794	33,839	1,105,349

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A18. Number of other school employees, by employee type and year

	Elementary Guidance Counselors	Secondary Guidance Counselors	Total Guidance Counselors	Library Specialist	Library Support Staff	Aides
1992-93	29,887	51,087	80,974	50,679	26,537	418,667
1993-94	31,761	51,478	83,239	51,666	26,414	449,845
1994-95	32,835	52,127	84,962	51,547	25,575	466,470
1995-96	34,348	53,533	87,881	51,788	25,896	488,815
1996-97	34,901	54,228	89,129	52,229	26,276	509,729
1997-98	36,169	54,752	90,921	53,021	26,796	546,931
1998-99	38,108	55,001	93,109	53,042	27,024	534,641

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

Table A19. New district measures in 1998-99, by region

	Northeast	South	Midwest	West	Territo-ries	Total
Diploma Recipients						
Total	421,129	839,414	631,212	540,953	33,208	2,465,916
Male	203,430	403,801	308,002	262,128	16,025	1,193,386
Female	217,699	435,613	323,210	278,825	17,183	1,272,530
Amer.Ind.	836	8,061	3,963	10,928	2	23,790
Asian	18,794	19,881	13,776	59,007	2,198	113,656
Black	47,403	185,143	57,234	28,207	990	318,977
Hispanic	30,286	82,854	19,506	116,506	29,971	279,123
White	323,810	543,475	536,733	326,305	47	1,730,370
Other HS Completers						
Male	83	12,981	431	2,916	5	16,416
Female	90	13,104	431	2,361	6	15,992
Amer.Ind.	1	63	2	136	0	202
Asian	3	286	24	786	11	1,110
Black	25	13,759	68	302	0	14,154
Hispanic	14	947	44	831	0	1,836
White	130	11,030	724	3,222	0	15,106
Limited English Proficient	549,218	883,515	308,918	1,829,844	326,395	3,897,890
Reduced Price Lunch	431,682	1,257,944	627,388	789,407	27,740	3,134,161
Total Free/ Reduced Price Lunch	2,654,511	7,188,517	3,083,617	4,278,755	580,223	17,785,623

Appendix B. Previous Longitudinal Editing and Imputation of CCD Data

Previous development of the Longitudinal Database was carried out in four rounds. In the initial round, no editing or imputation of CCD data was undertaken. This appendix describes the editing and imputation procedures used in rounds 2, 3, and 4. Some of the tables have been updated from original versions of the text for inclusion in this appendix.

Procedures used in the Second Round of File Development *(These procedures were previously described in McLaughlin, Huberman, Hawkins, and Lee, 1997.)*

The Common Core of Data relies on state-level aggregation of district information and transmission to NCES. In that process, there are occasions for errors in interpretation by respondents and errors of data entry. It is impossible to identify many errors because the resulting figures, by themselves, appear to be reasonable. However, when data from 8 years are merged, it is possible to make much more precise identification of errors. For example, a district whose reported enrollment pattern over 8 years is (375, 390, 365, 40, 415, 420, 410, 430) can be assumed to have a data entry error in the fourth year, and an enrollment of about 400 would be a reasonable estimate for that year. In preparing the CCD longitudinal report on small rural school districts, extensive editing and imputation were undertaken. The specific steps are described in this section. Chronologically, the 1986-87 through 1991-92 data were edited and imputed simultaneously, and the 1992-93 and 1993-94 data were subsequently imputed using the values from the preceding years. The editing and imputation was performed in the following 15 steps.

Step 1. Specify the records to be included. Identify school districts that change type from regular to nonregular and back, and set the type to be constant. Reported types of some districts in Maine, Massachusetts, California, Ohio, Virginia, and Vermont were changed in some years. (For one LEA on the Mississippi River whose state did not match its identification code, the variable STATE was changed.)   Also, if any district has no students, no teachers, and no schools, and does not merge with any schools on the school file, in any year, delete it from the file. This step determines the number of district records on each year's file.

Step 2. YEARS. Create YEARS, a string with one character for each year: "Y" if the district is on the district file and merges with at least one school on the school file in the year, "N" if the district is on the district file but merges with no schools on the school file in the year, and "M" if the district is not on the district file in the year.

Step 3. Number of schools. If the number of schools is missing for a district for a year, use the number from a preceding year with data. If the number is not available for any year, use the number of records on the school file for the district. (If none, set the number of schools to zero.)

Step 4. Grade span. If high grade and low grade are missing for a year, use the previous or closest year if some year has data. Otherwise, impute from school file. If the school file grade span is indeterminate, but there is a school, impute KG-to-12. Otherwise (if there is no school), impute as missing. Edit gradespans to remove cases in which low grade is higher than high grade and set them equal to whichever is not imputed, or if neither is, to the lower of the two.

Step 5. Number of teachers. Set spurious zeros for numbers of teachers (in Massachusetts and Michigan in 2 years) to missing. If number of teachers is missing in a district for a year, use the sum from the school file if there is a match. Otherwise, use a prior year's count, or if no teacher counts are available for any year, impute a value equal to the product of the number of schools times the number of grades in the gradespan (i.e., one teacher per school per grade). If the gradespan is indeterminate, impute one teacher per school.

Step 6. Edit number of students. Replace zero or missing values for enrollment in a district, or values that differ from an adjacent year by both 40 and 40 percent, with positive values from the school file whenever available. Note that when single years were added to the file later (i.e., 1992-93 and 1993-94), this step was repeated.

Step 7. Edit student/teacher ratio. Remove large or inconsistent student/teacher ratios (S/T). If for some year, a district's S/T is greater than 50 or S/T is inconsistent with both of the 2 adjacent years (by a factor of 2 or more), and the adjacent years are consistent with each other, then either set S to missing (to be imputed) or impute T directly. If S is consistent with adjacent years but T is not (each by a 40 percent factor), impute T as the average of the two years it is adjacent to. Otherwise set S to missing. One district, new in 1991-92, has number of teachers imputed from 1992-93, because its number of teachers in 1991-1992 created a student teacher ratio greater than 700.

Step 8. Impute number of students. Run PROC IMPUTE to impute total students in the 6 years. The imputation is by two categories of number of schools (districts with fewer than 4 schools and districts with 4 to 19 schools). No districts with more than 20 schools were missing total enrollment. The average number of schools and average number of teachers were used in PROC IMPUTE.

Step 9. Racial-ethnic percentages. This step imputes ethnic distributions. First, the SDDB (1990 decennial Census, mapped onto school district boundaries) is used to obtain percentages of each district's child population in different ethnic groups. For 27 districts for which no ethnic data are available for any year on the CCD or for the SDDB, impute the average for districts in the same city, or if not available, from the same county. For districts with data in some years but not others, perform the edit check described below, then use PROC IMPUTE. (However, no ethnic data were available for 1986-87, and none were imputed. Ethnic distributions for that year are not included in the report.)

Set inconsistent values to missing. These are values for districts that have values for at least 3 different years, and at least one of the percents differs from the average of all years by both (a) at least 25 percentage points and (b) at least 5 standard deviations. Also, for convenience, set the percentages for districts with zero students to the national averages: 1.1, 6.1, 5.4, 2.2, 85.2, for Asian, black non-Hispanic, Hispanic, American Indian/Alaska Native, and white non-Hispanic, respectively. Run PROC IMPUTE with the 20 variables (four ethnic groups (excluding white non-Hispanics) for each year from 1987-88 through 1991-92). An additional run using all years' data, but only imputing the last 2 years, was made to impute missing values for 1992-93 and 1993-94.

If the resulting sum of the minority percents is greater than 100 for any district, they are

normalized to 100. The white non-Hispanic percentage is set to 100 minus the sum of the other percentages in all districts.

Step 10. Locale code. For districts with schools with locale codes, the NCES standard procedure for deriving district locale codes from school locale codes was used. That procedure assigns the most frequent school locale code in the district, setting ties to the more urban locale, with the possible exception that for districts in which at least three-fourths of the schools have locales spread among values of 1, 2, 3, or 4 (i.e., in metropolitan areas) but the most frequent single school locale is 5, 6, or 7 (i.e., large or small town or rural), the district locale would be set to the most frequent of the values 1, 2, 3, or 4. (That exception did not occur in these data.)

For districts with no locale code in any year, the most frequent locale code for districts in the same county was used. If no data were available for the county, (a) the value 2 was imputed if the metro status code was 1; otherwise, if the number of schools was less than 5, the value 7 was imputed. If the metro status code was 2 and there were 5 or more schools, the value 3 was imputed; and if the metro status code was 3 and there were 5 or more schools, the value 6 was imputed. These rules are based on minimizing the percent errors based on relations observed for districts with data. Although the locale code was imputed separately by year, imputed values for a district were forced to be constant across years, equal either to the latest unimputed value or, if there were no unimputed values, to the modal value.

Step 11. Percent of school-aged children in poverty. (This variable was taken from the SDDB. It was therefore missing for all CCD districts not present in the SDDB.) The average percent poverty for districts in the same county was used to impute percent poverty. If there were no districts in a county with data, the average value 17 percent was used.

Step 12. Counts of special education students. First, counts in all districts in states which reported uniform zeroes in a year were set to missing, to be imputed. Second, if the number in a district exceeds the total number of students for a district, it was imputed to be equal to the total number of students.

Counts were then translated to fractions of total enrollment, and two variables were created: the average fractions for 1987-88 and 1988-89, and the average fractions for later years. Two averages were used because the values in the earlier years were not highly correlated with the values in later years. PROC IMPUTE was run, with five special education percentages (one for each year from 1987-88 through 1991-92), the two overall averages, and the percent of enrollment that was black non-Hispanic, plus American Indian/Alaska Native, minus Asian. It was run with separate hot deck distributions depending on whether there was a determined gradespan. These variables were selected on the basis of regression model results. Imputed percentages were translated back into counts.

Step 13. Four types of high school completers. Data were only available for the years after 1986-87, and the high school equivalence results were not available for 1991-92. First, values for 12th grade enrollment were imputed (and later dropped), in order to impute graduates as a ratio to the preceding year's 12th graders. Imputation of 12th grade enrollment occurred if the number of 12th graders was either missing, larger than the total enrollment, or less than half of the total completers

(the sum of four fields: regular diplomas, plus other diplomas, plus other high school completers, plus high school equivalencies).

If the grade span was reasonable, the value of the total enrollment divided by the number of grades was used for 12th grade enrollment. Otherwise, if there was a 12th grade and the number of completers was greater than zero, the grade 12 enrollment was set equal to the completers. If 12th grade was not offered or the number of completers was zero, count of 12th graders was imputed to be zero.

A small number of erroneous values for high grade in 1986-87 were set to 12. These were cases in which there were 12th graders enrolled and completers the next year but for which high grade was less than 12. Counts of completers were transformed to ratios to preceding years' 12th graders.

PROC IMPUTE was run after the file was prepared. Variables included were average ethnic percentages and percent in poverty, as well as the average over years of each of the four categories of completers. The latter averages, which normally would be no greater than 1, unless there was substantial in-migration, were not allowed to exceed 2. Values of percentage of 12th graders who earned regular diplomas that differed from the average (across years) by more than 50 percentage points and values of other completion types that differed by more than 20 percentage points from the average were set to missing. Hot deck distributions were selected separately for three sizes of 12th grade cohorts: <20, 20 to 99, and 100 or more. The results were transformed back to counts, and three districts new in 1991-92 were separately imputed to have no completers.

Step 14. All imputed counts on the file were rounded to integers.

Step 15. Impute Per-Pupil Revenues and Expenditures. In addition to variables on the CCD nonfiscal survey file, two variables on the F-33 Census of Governments survey, total revenues and expenditures per pupil, were imputed for the four school years from 1989-90 through 1992-3. For nearly every regular school district, data were present for at least one of the four years. Districts with data in none of these years (n = 90) were imputed as the average value of per-pupil revenues and expenditures for districts reporting data in the specified year, by category. The categories for which separate mean values were computed in each of the four years were large and small districts in rural and nonrural settings in each of the four standard geographic regions. (The division of the south into two subregions used elsewhere in this report was not applied to this imputation.)

For all imputations, the first step was to compute mean values of per-pupil revenues and expenditures for the 11,729 regular districts with F-33 data in all four of the school years (1989-90 through 1992-93). The mean values for per-pupil revenues and expenditures were obtained for each of four regions, separately for small and large rural and nonrural districts in each year (a total of 128 numbers). Means were weighted by the F-33 estimate of enrollment in the year.

Next, for each pair of adjacent years, a linear regression function was estimated, using a single predictor (the same measure in the adjacent year), to predict the deviation of a district=s per-pupil revenues or expenditures from the mean for that district=s region and size and locale category. A total of 12 regressions were estimated (3 pairs of adjacent years, in each order, for revenues and

expenditures). The regressions were weighted by the F-33 estimate of enrollment in the year being predicted. Then, for cases missing in a year, the value was imputed as the sum of (a) the mean value for the region by size by locale category for that year and (b) the estimated deviation from the mean based on the regression.

The percentages of data that were imputed for this report range from 0.0 percent to 47.7 percent, as shown in table B2. Except for race and special education counts in the earlier years, none of these percentages were as great as 20 percent. Although these percentages primarily represent missing data, some imputed values are the result of setting unreasonable reported values to missing. As a general rule, most imputed values were based on reported values for the same district in different years, using the rules summarized above. It should be noted that these percentages pertain only to regular school districts, as used in this report. Between 1,000 and 2,000 other entities are included in the Common Core of Data public school district release file.

Table B1. Percentages of values imputed on the district files used in round 2

				Year				
Variable	1986-87	1987-88	1988-89	1989-90	1990-91	1991-92	1992-93	1993-94
Small rural districts								
Gradespan	0.3	0.6	0.4	0.2	0.3	0.9	0.6	0.5
No. of Schools	0.0	0.3	0.2	0.1	0.0	0.0	0.0	0.0
No. of Teachers	3.1	2.8	11.8	0.6	2.6	3.2	8.0	3.9
No. of Students	21.7	0.5	0.3	0.2	0.5	0.4	0.3	0.2
Race (Low/High)	--	25.-40.	22.-34.	20.-28.	14.3	8.8	4.0	1.8
Special Ed Count	--	40.7	29.7	26.0	31.3	13.3	1.8	2.6
Locale	4.3	4.0	3.1	1.8	0.9	0.2	0.0	0.0
Per-Pupil Revenue	--	--	--	3.9	26.1	1.6	15.7	--
Per-Pupil Expenditure	--	--	--	3.9	26.2	1.5	15.8	--
All districts								
Gradespan	0.6	0.4	0.3	0.2	0.3	0.6	0.4	0.3
No. of Schools	0.1	0.2	0.1	0.1	0.0	0.0	0.0	0.0
No. of Teachers	6.4	7.7	13.9	2.6	5.8	6.6	5.7	2.1
No. of Students	19.4	0.5	0.2	0.1	0.3	0.2	0.2	0.1
Race (Low/High)	--	26.-36.	16.-25.	12.-18.	10.9	7.6	5.1	2.4
Special Ed Count	--	47.7	35.0	23.1	30.4	15.7	10.2	6.4
Locale	3.5	3.4	2.8	1.7	1.1	0.1	0.0	0.0
Per-Pupil Revenue	--	--	--	2.9	15.7	1.3	12.6	--
Per-Pupil Expenditure	--	--	--	2.9	15.7	1.3	12.6	--

Notes: -- Indicates that the measure was not included in this report for the particular year.
Percentages of race/ethnicity imputation, unlike other measures, are for schools.

Three of the entries for race/ethnicity in table B1 represent a range. Before 1990-91, there were different percentages of missing data for different race/ethnicities, ranging from a low for white non-Hispanics to a high for American Indians/Alaska Natives. District level race/ethnicity percentages were obtained by summing the percentages for schools in the district, with appropriate weights. However, there were a few districts with no school data. Therefore, in addition to the values imputed at the school level shown in table B2, small percentages of race/ethnicity distributions were imputed at the district level. These percentages were for 0.9, 1.0, 0.4, 0.5, 0.1, 0.2, and 0.3 percent of the districts in the years from 1987-88 through 1993-94, respectively.

Procedures used in the Third Round of File Development *(These procedures were previously described in McLaughlin, 1999.)*

The third round of file development consisted primarily of adding new years of district data to the existing longitudinal file. The procedures are described for adding the years 1994-95 and 1995-96, but they also describe the procedures used for adding 1996-97 and 1997-98 data in the fourth (current) round of file development. Tables have been revised to reflect the new addition of two years' data.

Phase I.

1.   The first step is to read the agency data and school data into SAS files and identify the extent of missing data. Names are assigned to the variables to conform to the names assigned for preceding years.

2.   The second step is to identify the set of districts to include in the longitudinal file. These include all districts also included in some preceding year, plus new districts. That is, the districts that had previously been excluded will continue to be excluded. Originally, only type 1, 2, and 3 districts were included, and that continues generally to be the case. However, when evidence indicates the need to include some districts previously excluded (e.g., regular districts found to consolidate into districts labeled as regional), they are added to the file. New districts are added to the file only if they are types 1, 2, or 3 and have at least one student enrolled.

3.   The third step is to identify each closure from the preceding year and to determine the district to whom most of its students probably transferred. For this, a printout of districts in the same state as a closing district, with a corresponding jump in enrollment, is examined. The printout is sorted by county and city, and an atlas is used to judge which other districts are sufficiently close to take the students. Grade spans are also taken into account in determining the "successor" district. Specifically, the district whose address is closest to the closing district, by road, among districts with a compatible grade range and enrollment, is identified. After the determinations are made, the variables NXTYRID and PRVYRID are assigned. During this step, closed districts that were on the file in the preceding year with zero enrollment for that year are reclassified as closed one year earlier and the successor districts are sought in the appropriate year.

4.   Next, the missing data variables are defined and preset and the variable YRS is extended on character to include the year to be added. The i-th character in YRS gives the status in year i, where i=1 corresponds to 1986-87. "Y" indicates that there is a district record and at least one school record for the district in the year; "N" indicates that there is a district record but no school record in the year; and "M" indicates that there is no district record in the year.

5.   The final step in Phase I is to create a file that merges all years' district data and a file that merges all years' school data, aggregated to the district level.

Phase II.

        Phase II consists of the imputation of the most basic information about school districts: the number of schools and the grade span.

1.   Although schools are opened and closed to respond to changing enrollment, changes of more than one school in a district need to be examined. The first step is to list the districts in which the number of schools changed by more than one but which did not experience a corresponding enrollment change.

2.   The enrollments and grade spans of the schools in these districts are compared in two adjacent years to determine whether there was a reorganization of grade levels in schools or a combining or splitting of schools that would explain the change in number of schools. There were very few cases which could not be explained in this way, and in most cases these appeared to involve counting some form of specialty school in one year and not in the other year. Numbers of schools were edited and imputed for a total of only 118 districts over the ten-year period (including 72 in 1986-87). Generally, explainable differences in numbers of schools (e.g., the closure of an alternative school) were left unedited, as were any changes of a single school in a district.

3.   Information on the numbers of schools was missing for 66 districts over the ten-year period, and these were imputed to be the same as in an adjacent year. In the few (8) districts which were on the file for only one year, a total of eight schools were imputed.

4.   Most grade span changes from one year to the next involve prekindergarten, kindergarten, and first grade. These were not edited unless they were in conflict with information from the school file. With the exception of 1986-87, few reported grade spans were replaced, and most of those were from "00" to a legitimate value, possibly "UG" (ungraded). Values of "00" were generally changed to values from an adjacent year's district file. In 1986-87, there appears to have been an excess of reported prekindergartens, compared to all later years.

5.   Grade spans in schools with more than 30 students, where there was a change not involving PK, KG, or 1, were considered for replacement. They were replaced if they constituted a "V" (that

is, a change that was reversed in the following year) with no change in enrollment that would correspond to the addition or removal of a grade.

1.   For 1995-96, a few new districts reporting "00" for grade spans were compared with grade spans reported on a State Education Agency web page. The grade spans reported on these web pages were substituted for the "00."

Numbers of cases imputed are shown in table B2, (a) for fields missing in the original CCD and (b) for values judged erroneous. Although the text in this section is based on Round 3 imputations, the imputation counts in tables B2 through B4 represent cumulative results across all four rounds of imputation.

Phase III.

Phase III consists of imputing the two basic counts for each district: the numbers of students (MEMBER) and of full-time equivalent teachers (FTE). These two counts provide the context for imputing other counts, such as the numbers of graduates, ungraded students, and special education students. These two measures must be edited in a coordinated fashion, because the primary criterion for validation of the data is the student/teacher ratio. If the ratio is out of a reasonable range, it is necessary to decide (based for example on another year's data) which of the two numbers is in error.

1.   The first step is to replace missing and zero teacher counts on the district file with counts aggregated from the corresponding school file records, if the resulting student/teacher ratio would be between 1 and 100.

2.   Student/teacher ratios of greater than 100 are also identified for potential replacement of either the student count or the teacher count.

3.   First, teacher counts are considered for replacement. Six linear relations are estimated, three for districts with more than 500 reported students and three for smaller districts: (a) predicting the teacher count from the student count; (b) predicting the student/teacher ratio from the prior year's ratio; and (c) predicting the prior year's student/teacher ratio from the year preceding that. School districts with identified potential problems with these counts are excluded from the regressions.

4.   If there is information from the preceding year and the teacher count changes by more, in terms of percentage, than the student count, the teacher count is imputed by dividing the student count by the regression estimate of the student/teacher ratio based on the previous year. The regression estimate of the student/teacher ratio includes random error as indicated by the regression estimation procedure. Imputed teacher counts are limited to be larger than 1/100 of the corresponding student counts.

Table B2a   Number of missing values for which imputations were generated: Part 1

	SCH	GSLO	GSHI	MEMBER	FTE [1]
1986-87	9	10	10	2,936	15,360
1987-88	18	10	10	14	1,166
1988-89	9	0	0	8	2,060
1989-90	8	0	0	2	383
1990-91	1	1	1	1	1
1991-92	0	36	36	0	924
1992-93	0	0	0	1	550
1993-94	0	0	0	0	269
1994-95	0	0	0	2	137
1995-96	2	0	0	3	387
1996-97	6	0	0	639 [2]	848 [2]
1997-98	0	0	0	2	524
1998-99	0	6	6	31	210

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File. [1] Summed from school file in 1986-87. [2] Originally reported values missing in New Jersey in 1996-97.

Table B2b. Number of non-missing values for which replacement imputations were generated: Part 1

	SCH	GSLO	GSHI	MEMBER	FTE
1986-87	762	2,294 [1]	82	489	0
1987-88	48	49	44	145	63
1988-89	43	44	45	46	85
1989-90	65	46	48	26	18
1990-91	53	64	60	49	956
1991-92	38	23	23	544	60
1992-93	98	47	41	24	372
1993-94	29	40	31	17	212
1994-95	29	8	8	16	184
1995-96	34	63	67	61	127
1996-97	27	89	86	931	125
1997-98	10	23	23	920	159
1998-99	15	113	100	51	171

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File. [1] In 1986-87, 2,207 districts that reported "PK" as the lowest grade were set to "KG" to match 1987-88 figures and school figures.

5.   If there is no preceding year's information, teacher counts are imputed using the regression based on students, including the addition of random error with variance indicated by the regression estimation. Imputed teacher counts are again limited to be larger than 1/100 of the corresponding student counts.

6.   The next step is to identify "Vs" in the student and teacher counts for the year preceding the most current year. A "V" is defined as a district value for teacher or student counts in a year which differs by a ratio of more than 7:5 or 5:7 (and by more than an absolute count of 40) in the same direction from both of the adjacent years, accompanied by a corresponding change in the student/teacher ratio by more than 20 percent.

7.   Teacher counts for "Vs" are imputed for the preceding year, based on the regression estimate of the student/teacher ratio from year preceding that. That is, the method is the same as in Step 4, but for one year earlier.

8.   Steps 1 through 7 are repeated for student counts, reversing the role of teacher and student counts. In repeating Step 3, separate regressions are computed for large and small schools, where the size criterion is 50 teachers.

9.   If both student and teacher counts are missing, the prior year's counts are used. This occurred in no cases in 1994-95 and one case in 1995-96.

10.   Finally, the student/teacher ratio is computed and correlated across adjacent years as a check on the adequacy of the editing. If the correlation is less than .85, there are probably additional problems with these counts that would interfere with use for longitudinal analyses. (The minimum value for the ten-year period was .87, between 1991-92 and 1992-93.)

Phase IV

In the next phase, information is added to the district file from the school file and the 1990 Census School District Data Book (SDDB). In particular, the counts of five racial/ethnic categories of students and free lunch eligible students were imputed for the years in which those measures were collected; district locale code was imputed from schools but held constant for the years 1986-87 to 1991-92 for the purpose of particular longitudinal analyses; and the percentage of school-aged children in poverty was imputed for districts not in existence in 1990 or otherwise not in the SDDB.

1.   Racial/ethnic counts are available for individual schools but not for districts as a whole. However, it is straightforward to aggregate the data from schools to the district level, and this is the first step in this phase. Counts of Asian/Pacific Islanders, Blacks (non-Hispanic), Hispanics, Alaskan Natives/American Indians, and Whites in enrollment were computed for each district by summing the counts for schools. Starting in 1990-91, racial/ethnic counts were reported for either none or all of the five categories, and prior to that, partially missing data could reasonably be

assumed to represent zeros. Therefore, the five counts are edited and imputed as a unit, with a single missing data indicator.

2.	The five counts were temporarily transformed to fractions of the total sum of the five counts for the purposes of editing and imputation.

3	Whenever there are missing data for districts which report the racial/ethnic counts in other years, linear regressions predicting percentages based on adjacent years have high accuracy rates (r^2>.96). These regressions are used for predicting the five percentages, and they are normalized to sum to 100 by dividing by their sum. A random disturbance is added to each regression estimate, matching the error variance indicated by the regression printouts.

4.	In a few cases, there are no data from other years. In these cases, the 1990 Census School District Data Book estimates of percentages of children by racial/ethnic category for districts in the same city or county are used. In these cases, it is necessary to verify that the address on the CCD file is the location of the district's school(s), not a regional central office.

5.	A few "Vs" in racial/ethnic counts were identified. The criterion used is a discrepancy of more than 15 percent in any of the five percentages from each of the adjacent years, which themselves are within 5 percent of each other. This criterion is limited to school districts with more than 300 students. Identification of "Vs" in smaller districts would not be reliable.

6.	When "Vs" are found, values are imputed, as in Step 2, from the prior year; that is, from the year before the "V."

7.	As a check on the imputations, the correlations of reported percentages between adjacent years are compared to the correlations of imputed percentages. As a final step, the racial/ethnic percentages are translated into counts, adding exactly to the total enrollment in the district, even though the reported race-ethnic counts and total enrollment counts may have been taken on different dates.

8.	Locale codes are aggregated from school locales by identifying the locale code that is the mode for schools in the district. In the case of ties, the more urban of the choices is assigned. These are not counted as imputed.

9.	Because the file was at one point developed in the context of a project that called for setting a single locale code for the period from 1986-87 through 1991-92, the values for this period are all set to be the same. That is not true for years from 1992-93 on, however. For 1992-93 and 1993-94, the school data were given priority over the previous years' data in computing the district locale code; and for 1994-95 and 1995-96, the previous year's data were given priority over the school data. That is, the school data would only be used if there were no prior year's data (i.e., it was a new district).

10.   For new districts for which a locale could not be derived from the school file, the locale of other districts in the same city were considered, along with information about the population of the city of the district. There were no such districts in 1993-94, but two in 1994-95.

11.   For percent poverty, the1990 Census data were used for all school districts in existence in 1989-90. For new districts in 1994-95 and 1995-96, a linear regression was used to estimate the percent of school-aged children in poverty. The regression made use of several predictors which were found to be significantly correlated with the percent poverty: the state mean percent poverty, the district aggregated percentage of students eligible for the Federal school lunch program, percent minority enrollment, student/teacher ratio, and whether the locale of the district was in a central city of a SMSA (locale=1 or 3). The r^2 for the regression was .54, so there was a substantial random component in the few cases imputed.

12.   The count of students eligible for the Federal school lunch program (FLE) was first aggregated from school-level data. In the earlier years, large percentages of the data were missing or reported as uniform zeros in a state. It was necessary to identify the year in which each state began to report positive numbers.

13.   Eleven states were identified for which no FLE counts were reported in any year. Nineteen other states failed to report counts in one or more years during the period. A series of zeros prior to the first non-zero value reported by a district was judged erroneous if the first two non-zero values were within a ratio of 3:4 of each other.

14.   Two initial linear regressions were run to remove outliers and create a database on which regression coefficients for imputations could be estimated. (Outliers were defined as any free lunch eligible percentages of enrollment that were more than 30 percent from the linear trend estimate, plus any zeros for which the linear trend estimate was greater than 20 percent.) The dependent measure for the regressions was the percentage of enrollment that was FLE, and when data from other years were available, the predictor was the interpolation or extrapolation of the linear trend estimate for that district. The r^2 was .96.

15.   When no data were available for any year, a more complex regression was used. For districts reporting F33 financial data in 1992, the revenue for the federal school lunch program, divided by the total number of student, was used, along with three other F33 measures, percent poverty, percent minority, percent special education students, and three categories of enrollment size (less than 30, between 30 and 500, and greater than 500). The r^2 for this regression was .78. For cases without F33 data, the r^2 was .60.

16.   As a check on the imputations, the overall trend in total national counts was examined, and mean imputed values were compared to mean reported values for the same districts in different years.

Table B3a. Number of missing values for which imputations were generated: Part 2

	ETHNIC	FLE	LOCALE	PK12	UG
1986-87			15,346		
1987-88	3,467	2,862	20	23	3,743
1988-89	2,367	3,046	13	8	3,358
1989-90	1,990	7,947	12	2	3,202
1990-91	1,405	8,667	22	1	1
1991-92	868	5,412	13	1	3,437
1992-93	553	3,990	4	1	3,326
1993-94	135	3,079	4	0	4,322
1994-95	118	2,823	4	2	3,663
1995-96	171	3,090	9	3	3,749
1996-97	773 [1]	2,732	44	641 [1]	4,372 [1]
1997-98	136	2,750	45	3	3,937
1998-99	149	2,986	10	28	4,259

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File. [1] Originally reported values missing in New Jersey in 1996-97.

Table B3b. Number of non-missing values for which replacement imputations were generated: Part 2

	ETHNIC	FLE	LOCALE	PK12	UG
1986-87			0		
1987-88	53	8,482	0	1,099 [1]	735
1988-89	142	7,609	0	971	52
1989-90	87	1,579	0	956	737
1990-91	84	929	251	975	934
1991-92	82	401	502	1,451	915
1992-93	85	393	78	934	919
1993-94	65	266	458	922	912
1994-95	47	140	0	924	911
1995-96	61	187	32	964	904
1996-97	44	3	0	931	10
1997-98	59	6	0	920	0
1998-99	502	37	0	49	4

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File. [1] Revised special education counts in Illinois led to reestimation of numbers of ungraded students.

Phase V.

The final phase of the editing imputation focused on subsets of total enrollment: students in grades prekindergarten through 12 versus ungraded students, regular and other diploma graduates and other high school completers, and special education/IEP students.

1.   For the division of enrollment between ungraded and other students, a simple imputation was carried out, either setting the ungraded count to zero, if the students in grades was equal to the total enrollment, or using the prior year's breakdown. Almost all ungraded imputations were zero.

2.   Special education counts were imputed using a regression predicting the percentage of students in special education from the state mean percentage, the percentage in the prior year, and the student/teacher ratio, resulting in an r^2 of .59. An indicator of whether the district had any ungraded students was also included but failed to contribute significantly to the prediction.

3.   "Vs" in special education counts were identified using a criterion factor of two; that is the count being considered must be more than twice or less than half of both the adjacent years' counts, with a discrepancy of at least 40 to guard against large percentage changes in very small districts. "Vs" were replaced by selecting a random value from a distribution whose mean was the average of the two adjacent years, and whose standard deviation was equal to the difference between those two values.

4.   The three completion counts (regular and other diplomas and other completers) were all edited and imputed in the same manner. The first step was to compute the district count of twelfth graders in the preceding year by aggregating records on the school file, to use as a denominator for three completion rates. Zero was imputed for missing rates if there were no twelfth graders reported in the preceding year and the school was not ungraded. State policies in granting diplomas were taken into account to the extent that no OTHHSC recipients were imputed for states that did not recognize this credential, according to the states' reports on the 1996-97 CCD.

5.   Numbers of graduates or completers were considered unreasonable if they were greater than the sum of the number of prior year's twelfth graders and the square root of the prior year's twelfth graders. That is, if there were 100 twelfth graders the prior year, more than 110 graduates from that class was considered unreasonable and replaced with that value.

6.   Linear regressions were used for imputation of high school completion rates, using as predictors the prior year's rate, plus percents minority, special education, and poverty (Census). The r^2 values were modest (.25, .48, and .30 for regular diploma, other diploma, and other completer rates), but the standard deviations around the mean values of .89, .01, and .05 were sufficiently small to justify the imputation.

7.   Finally, a single SAS file containing all of the quantitative data (ALL.SD2) was created, along with ten single-year files (AIRLEAyy.SD2, where yy=86, ... , 95), each containing directory and other information copied from the CCD files. Missing data indicators take on the value "M" for imputed data and a single blank character for reported data.

Table B4a.   Number of missing values for which imputations were generated: Part 3

	REGDPL	OTHDPL	OTHHSC	HSEQVR	SPECED
1986-87					
1987-88	2,777	9,813	10,021	10,150	4,286
1988-89	1,936	9,130	9,364	9,387	2,032
1989-90	2,129	8,838	8,553	9,435	1,725
1990-91	1	1	1	1	1
1991-92	350	8,776	10,016		1,398
1992-93	3,885	10,035	13,310		860
1993-94	3,750	10,732	13,195		256
1994-95	3,695	10,156	13,091		450
1995-96	3,432	11,034	13,202		354
1996-97	3,791	11,096	13.296		1,067 [1]
1997-98	4,067	11,264	13,185		199
1998-99					126

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File. [1] Originally reported values missing in New Jersey in 1996-97.

Table B4b.   Number of non-missing values for which replacement imputations were generated: Part 3

	REGDPL	OTHDPL	OTHHSC	HSEQVR	SPECED
1986-87					
1987-88	644	17	11	39	3,517
1988-89	470	28	16	35	3,765
1989-90	456	32	21	41	2,460
1990-91	819	1,450	797	5,267	5,308
1991-92	429	45	38		1,795
1992-93	256	78	23		2,766
1993-94	296	64	36		1,560
1994-95	16	3	3		964
1995-96	12	1	0		1,563
1996-97	4	2	0		39
1997-98	0	0	0		33
1998-99					43

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

As a result of selecting a subset of districts for inclusion in the Longitudinal CCD Non-Fiscal Survey Database and of editing and imputing values for some of the fields, the total counts of quantitative statistics of public elementary and secondary education are somewhat different from values based on (a) the unedited counts in the original (full) set of CCD districts and (b) the unedited counts in the longitudinal subset of districts. The differences for fields reported on the CCD district survey are shown in tables B5, B6, and B7.

The largest differences in total students and teachers (table B5) are in 1986-87, when no student data were reported by 10 states and no teacher data were reported at all. (Counts were reported on the school files in that year, however, and these can be aggregated to produce total count estimates.)

The largest differences in ungraded and special education counts (table B6) are that the longitudinal subset omits local education agencies serving special populations, who in many cases are in ungraded settings, and that special education counts were only partially reported in the earlier years studied.

The largest differences in diploma and completer counts (table B7) are that the imputed counts are higher. This arose because many districts which reported other diplomas or other completers in some years left this field blank (not zero) in other years. Counts comparable to years in which they reported counts were imputed in the years in which they left these field blank. If, in fact, the missing data for these two counts actually reflect zero values (i.e., no other diplomas or other completers), then the longitudinal file may overestimate the actual totals for other diplomas and other completers.

Table B5. Comparison of school, student, and teacher counts based on edited and originally reported values

	Number of schools			Number of students			Number of teachers		
	Original file	Longitudi nal subset	Imputed	Original file	Longitudinal subset	Imputed	Original file	Longitudi- nal subset	Imputed
1986-87	84,755	81,875	84,249	31,798,484	31,310,928	40,348,614	.	.	2,222,912
1987-88	85,063	82,235	84,141	40,706,279	39,747,330	40,563,711	2,106,816	2,047,606	2,269,376
1988-89	84,911	82,195	84,039	41,039,846	39,933,991	40,712,087	1,978,032	1,913,772	2,299,633
1989-90	85,156	82,462	84,400	41,453,526	40,335,422	41,052,724	2,331,819	2,264,938	2,357,993
1990-91	86,277	83,469	85,339	42,095,467	41,000,579	41,726,161	2,286,589	2,218,535	2,395,186
1991-92	86,287	85,166	85,314	42,767,578	42,561,580	42,534,044	2,297,463	2,273,261	2,418,786
1992-93	86,089	85,083	85,387	43,436,788	43,279,584	43,290,970	2,396,342	2,362,613	2,450,164
1993-94	87,104	86,049	86,073	44,077,650	43,914,555	43,917,655	2,533,470	2,499,195	2,511,789
1994-95	88,099	86,757	86,805	44,777,473	44,527,708	44,532,250	2,524,181	2,486,806	2,555,458
1995-96	88,981	87,582	87,634	45,495,501	45,242,354	45,262,779	2,594,806	2,555,947	2,603,306
1996-97	90,070	88,616	88,710	45,312,274	45,053,747	46,030,930	2,547,896	2,509,500	2,665,632
1997-98	91,340	89,773	89,804	47,019,550	46,767,433	46,518,775	2,656,508	2,618,037	2,740,294
1998-99	92,884	91,286	91,292	47,125,496	46,887,400	46,979,705	2,785,035	2,745,732	2,830,163

Table B6. Comparison of ungraded and special education student counts based on edited and originally reported values

	Number of ungraded students			Number of special education students		
	Original file	Longitudinal subset	Imputed	Original file	Longitudinal subset	Imputed
1987-88	801,609	697,071	869,610	1,955,846	1,922,545	3,977,793
1988-89	916,238	620,578	803,238	2,503,359	2,471,028	4,111,821
1989-90	842,641	726,349	904,226	3,406,534	3,329,000	4,278,440
1990-91	794,402	680,870	858,703	3,285,855	3,211,855	4,393,260
1991-92	858,347	762,855	930,021	3,734,509	3,696,668	4,503,128
1992-93	782,982	709,553	881,439	4,058,204	4,007,990	4,808,173
1993-94	795,104	723,221	897,439	4,550,921	4,493,412	4,939,141
1994-95	761,399	689,972	855,654	4,604,596	4,541,355	4,862,792
1995-96	765,390	695,186	870,866	4,552,232	4,488,737	4,921,007
1996-97	864,690	810,829	883,353	5,045,138	4,987,440	5,131,819
1997-98	953,306	889,315	889,319	5,568,445	5,506,711	5,484,770
1998-99	811,148	764,418	628,114	5,765,190	5,701,702	5,685,413

Table B7. Comparison of regular and other diploma and high school completer counts based on edited and originally reported values

	Number of regular diplomas			Number of other diplomas			Number of other high school completers		
	Original file	Longitudi-nal subset	Imputed	Original file	Longitudinal subset	Imputed	Original file	Longitudi-nal subset	Imputed
1987-88	2,477,492	2,435,488	2,335,291	23,474	17,883	31,451	6,087	5,938	8,444
1988-89	2,462,473	2,417,929	2,511,651	24,225	19,764	31,753	9,866	9,663	17,732
1989-90	2,385,885	2,333,313	2,450,973	27,808	23,653	28,760	13,048	12,984	14,302
1990-91	2,277,010	2,224,643	2,332,513	26,234	21,176	44,790	14,935	14,813	18,869
1991-92	2,255,354	2,242,927	2,244,970	35,745	35,664	36,382	14,119	13,889	15,072
1992-93	2,188,933	2,177,245	2,210,838	63,936	63,816	71,536	17,841	17,482	23,415
1993-94	2,216,376	2,204,868	2,247,646	31,604	31,317	39,421	24,865	24,562	30,689
1994-95	2,177,083	2,166,045	2,226,577	34,717	34,574	37,866	26,123	25,891	28,087
1995-96	2,294,627	2,280,149	2,285,110	33,059	32,989	41,045	27,735	27,473	36,434
1996-97	2,241,204	2,226,438	2,292,086	31,029	30,931	37,464	31,704	31,397	43,266
1997-98	2,272,292	2,258,307	2,360,948	32,370	32,326	40,014	39,301	39,116	48,069
1998-99	2,478,851	2,464,800	2,465,916				32,666	32,408	32,408

Note: In 1998-99, regular and other diploma recipients were combined.

Procedures used in the Fourth Round of File Development *(These procedures were previously described in McLaughlin, 2001.)*

As a first step in creating longitudinal files, a correct matching of records from year-to-year is necessary. For the school file, this involves both matching schools to districts and linking schools that changed districts so that the fact that they were the same school can be used for longitudinal analyses. For the district file, this involves examination of each closure to determine which district(s) served the students in the following year.

Matching the number of records on the school file to the number of schools recorded on the district file. This is not the straightforward computational task that it might seem (i.e., just adding up the number of records on the school file and putting that number on the district file) because the presence of a discrepancy is a valuable indicator that some error in reporting has occurred. That error is likely to be the misclassification of a school in the wrong district, multiple records for a school, or an error in recording that the school is open or closed. Analysis of discrepancies can clarify the change that is needed in either the school or district file to increase the accuracy of the combined database.

1.   The starting point was a set of thirteen individual year files on which Synectics, Inc., had carried out preliminary editing. NCES assigns a 12-digit code number to each public school (2 digits for the state FIPS code, 5 digits for districts within state, and 5 digits for individual schools. In most cases, the 5-digit individual school codes are intended to be unique within a state. Because many schools remain in continuous operation while the districts to which they are assigned are changed, the same schools frequently have different NCES school codes in different years. To identify continuing schools and to differentiate them from openings and closings of schools, Synectics, Inc., assigned a preliminary "Master ID" for each school, equal to the NCES code it was given in its first year on the file. Between 1 percent and 2 percent of schools changed NCES codes in the period from 1986-87 to 1997-98.

2.   The first step was to verify the NCES ID code changes found by Synectics, Inc., and then add to these. This was done in 12 sub-steps, starting with 1997-98 data and ending with 1986-87 data. In each year, the list of districts with non-matching numbers of schools (between district and school files) was prepared, and the schools in the districts on that list were examined manually to identify the source of each discrepancy. For each year, (a) duplicate records for the same school were identified and deleted; (b) schools found to match schools with different Master IDs in the preceding year were reassigned the appropriate Master ID; and (c) the number of schools recorded on the district file was altered if needed. Master Ids were changed for approximately 2,000 schools in this process. The changes are codified in a SAS program written (a) to produce printouts for examination and (b) to implement changes to the files. The program commands in that program are idiosyncratic to each year, addressing editing contexts particular to each year. A typical example of the hundreds of segments of code in the program is the following, taken from the step to edit the 1987-88 school file.

```
if masterid="050810000536" then do;
                          masterid="050810001403";
                          ncessch ="050810001403";
                    end;
```

That is, the school which had been assigned the Master ID of 050810000536, was reassigned the code of 050810001403, which was the number assigned to that school (Humphrey High School in Humphrey, Arkansas) in later years. In that case, both the Master ID generated by Synectics, Inc., and the NCES school code were changed. In other cases, other changes were made.

3.   At the completion of the preceding step, a school file was created for each year with numbers of records in each district matching the number of schools recorded on the longitudinal district file for that district in that year. Table B8 illustrates the kind of information that can be derived from the edited files. The examinations in step 2 were limited, however, to those districts in which the numbers of schools initially failed to match the number recorded on the district file. There was no assurance that no schools were misidentified in districts in which the total numbers matched. Thus, in the following step of the school file editing process (editing of enrollment counts), it turned out that a large percentage of the discrepancies between school and district enrollment counts were not errors of counting students but rather errors in identifying the districts in which the school's students were counted.

Table B8. Number of schools opening and closing, by year

Year of New Status	Open	Close
1987-88	1,390	1,493
1988-89	1,219	1,301
1989-90	1,558	1,186
1990-91	2,105	1,146
1991-92	1,502	1,521
1992-93	1,734	1,670
1993-94	1,750	1,068
1994-95	1,715	1,007
1995-96	1,757	891
1996-97	2,100	1,019
1997-98	1,937	892
1998-99	2,778	1,310

Matching district closures/consolidations across years. Information on each school district that disappeared from the CCD file (i.e., closed) after some year during the ten-year period was examined to determine the most likely receiver of its students. Generally, a geographically close district, with the appropriate grade span and exhibiting a matching increase in students the following year, was identified as the receiving district. In some cases, the year of closing was not the same as the year in which the record was removed or the year preceding this removal. An enrollment of zero students was taken in some, but not all, cases as an indication of which year the district closed. Fairly clear identifications were possible for nearly all of the districts that enrolled 25 students or more the year before they closed; however, receiver districts for the very small district closures, many of which were in Nebraska, are ambiguous.[5] Undoubtedly, when districts closed, some students enrolled in different districts, some moved, some attended private schools, and some dropped out. Thus, when very small districts closed, the effects on the enrollment of nearby districts were invisible.   Therefore, the numbers of school district closures shown in table B9 are considered estimates.

Table B9. Number of district closures/consolidations, by region and year

	Northeast	South	Midwest	West	Total
1987-88	14	19	62	16	111
1988-89	16	5	56	17	94
1989-90	8	16	60	24	108
1990-91	3	24	62	22	111
1991-92	13	32	71	14	130
1992-93	11	21	111	34	177
1993-94	12	33	116	52	213
1994-95	28	11	70	46	155
1995-96	5	4	69	26	104
1996-97	9	3	33	23	68
1997-98	5	11	33	40	89
1998-99	4	2	93	19	118

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District File.

The variable NXTYRID is the 7-digit NCES ID code for the district that appeared to inherit most of the students from a closing district. An inverse variable, PRVYRID, was added to the receiving district on next year's file. PRVYRID is the 7-digit NCES ID code for the closing school district that sent most students to this district. (In a few cases, such as reorganizations, multiple

---

[5]   In addition to having many small districts, Nebraska is also problematic in that consolidation may join several districts that are not geographically contiguous.

districts closed and a single district inherited all of their students.) It is important to note that PRVYRID only indicates inheritance of students from *closing* districts. Reorganizations that did not result in removal of a district from the CCD universe are not identified. In particular, the addition of charter schools as districts on the CCD file implies transfer of students from public school districts that remain in operation, but these linkages are not identified.

Editing and imputation steps for the developing the Longitudinal School File

During the fourth phase of the development of the CCD Longitudinal Nonfiscal Survey Database, twelve years of responses to a school-level survey were added to the database. For each public school, information is available on the basic CCD school files since 1986-87 on the number of enrolled students, overall and by grade, and since 1987-88, also on the number of full-time equivalent teachers, race distributions of enrollment, and free lunch eligibility counts. The first two steps in incorporating school-level information were time-consuming:

(1)   matching the number of records in the school survey in each district to the number of schools reported on the district survey, by editing NCES school codes and identifying the schools whose NCES school codes changed between years, and

(2)   editing enrollment and full-time equivalent teacher counts to eliminate those mismatches between each district's counts and the sum of counts in schools in the district that are due to reporting errors (without eliminating plausible mismatches attributable, for example, to non-school programs for some students).

The editing of enrollment, teacher, and race/ethnicity counts were each accomplished as a series of computations.

Matching the sum of enrollments in schools in a district to the enrollment total recorded on the district file. First, it must be realized that these numbers need not match exactly. Many regular school districts serve students who are not assigned to a particular school. Usually the discrepancies are small as a percentage of a district's enrollment, but their existence means that an exact match cannot be taken as the criterion for accuracy of the file. Therefore, at the conclusion of the editing and imputation, differences remain in the total enrollments based on school and district surveys (compare tables A3 and A13).

1.   As a starting point a longitudinal school file was created, containing enrollment and FTE teacher counts for each year from 1986-87 through 1997-98. The result of editing that file was the creation of the file SCH12YRS (whose contents are listed in Appendix C). Records on that file are grouped by the district they are in, so that a school that was in two districts in different years appears twice on the file, once associated with each district. To distinguish multiple occurrences of a school, each record contains a 12-character variable, LYRS. "Y" in the i-th position indicates that the school was in the district indicated in the LEAID field of the record in the i-th year (1986-87=1), and "N" indicates that it was not. A second variable, SYRS, indicates all the years in which the school was open ("Y" in the i-th position indicates *open*, "N" indicates *closed*). For example, a school that was open in 1990-91 but changed districts in 1994-95 would have "Y" in position 5 of

SYRS on both records, but position 5 of LYRS would be "N" for when it appeared in one district and "Y" when it appeared in the other. In the following hypothetical example, Washington School, which opened in 1987-88, has a record for Smithville Elementary District and a second record for Smithville Unified District, created by a consolidation in 1994-95.

(hypothetical example)	Smithville Elementary District	Smithville Unified District
Washington School: SYRS	NYYYYYYYYYYY	NYYYYYYYYYYY
Washington School: LYRS	NYYYYYYYNNNN	NNNNNNNNYYYY

This file structure is particularly useful for imputing missing enrollment data because information can be used from adjacent years, even if the school was in a different district. In fact, most jumps in district enrollment figures from one year to the next are explainable as the change of a school to a different district.

2.　　A preliminary categorization of types of discrepancies was carried out to identify cases requiring manual examination because they fit no systematic pattern. Enrollments in schools in 148 district-by-year combinations were analyzed manually, and school and district files were edited as needed to minimize discrepancies. This analysis made clear that most enrollment discrepancies between school and district files were the result either of misassignment of schools to districts on the file or delay in adding a school to the file.

3.　　A criterion was set for identifying discrepancies. A discrepancy is either a missing value or a difference between school and district enrollment figures of greater than 10 percent of the higher value, but at least 50 students. Thus, if the sum of school enrollments was 89 and the district enrollment was 100, that discrepancy would not be examined, but if the two numbers were 899 and 1000, the discrepancy would be examined.

4.　　More than one-third of the discrepancies were associated with one state in two years: Illinois in 1996-97 and 1997-98. In those years, many special education students were double-counted in district reports. Discussion with the Illinois CCD coordinator clarified the situation, and these discrepancies were removed. The Illinois CCD coordinator also provided figures on special education counts for the entire period of the longitudinal file, and the Longitudinal District File was amended to reflect this information.

5.　　Next, 206 discrepancies in which the school figures were clearly more credible than the district figures were identified, and the Longitudinal District File was amended accordingly. Credibility was evaluated in terms of compatibility with adjacent years' figures.

6.　　Next, 160 discrepancies which could be explained by a change in one school's enrollment to be the same as its value in an adjacent year were identified, and those school enrollments were amended accordingly.

7.　　Next, 1,152 discrepancies that could be explained as "V's" (i.e., large one-year deviations from preceding and following years' figures in one field, such as enrollment counts, not corroborated by a corresponding deviation in another field, such as teacher counts) were examined,

and those school enrollments were amended accordingly. Also, 1996-97 figures for New Jersey were imputed as averages of adjacent years.[6] Note: in each of these steps, one or more special cases were uncovered, and enrollment figures were edited as appropriate.

8.   At this point, 804 discrepancies remained. A series of global steps was undertaken to reduce this to an acceptable number. Although these steps did not involve case-by-case examination of the figures, diagnostics were examined to ensure that the results were plausible. The first of these steps was to impute backward in time. If a discrepancy could be explained by assuming a school had really been open a year earlier than it appeared on the CCD file, with an enrollment similar to that in the following year, the school was "added" to the file in the preceding year. The second step was an analogous forward imputation from one year to the next. These two steps were especially effective in filling in the cases in which a school's enrollment was not reported at the school level during a year in which a district was reorganizing, although those students continued to be reported at the district level. (It is highly unlikely that those students were not in some school in the district in the intervening year.) The criterion for applying this step was the following. First, the district enrollments in the two adjacent years were required to be within 10 percent of each other; second, the school-to-district match had to be within 5 percent in the "good" year; and third, 75 percent of the discrepancy had to be explained by either missing school enrollments or school enrollments that were lower than the adjacent year by a factor of at least two. Also, there must have been some schools in the district that were not counted in the third part.

9.   In the same manner, schools were removed from the file a year earlier than they were reported to have closed when that removed a discrepancy. These appeared to be cases in which two schools counted the same students, perhaps because a new school opened after the beginning of the school-year. These examples suggest the need for a refinement in the enrollment counts: counting "FTE" students who attend one school full-time for one school-year. Using such a measure, districts might address the issue of students' attendance at two schools in a year by counting them as half in each school.

10.   The two steps described in (8) above were then repeated, with a more lenient definition of "explains the discrepancy." The district enrollment counts in adjacent years were allowed to be 30 percent different, and the schools with missing or low enrollments were only required to account for 60 percent of the discrepancy.

11.   At this point a list of the 15 remaining cases with discrepancies of more than 1,000 students was examined, on a case by case basis, and imputations were made as appropriate to remove the discrepancies.

12.   The next step was to add the territorial data for American Samoa, Guam, the Marianas, Puerto Rico, and the Virgin Islands. Note that the original CCD FIPS codes for the territories were changed between 1990-91 and 1991-92. The newer values are used in the longitudinal files.

---

[6] New Jersey did not report figures to CCD in 1996-97. NCES has prepared a revised 1996-97 CCD file, substituting 1995-96 figures for the missing 1996-97 data in New Jersey.

13. An examination of overall state sums of enrollment indicated three anomalies that could not be attributed to particular schools or districts. In Tennessee, 1986 district enrollments were about 30,000 greater than both (a) 1986 school enrollments and (b) 1987 district and school enrollments. In most years, the annual change in enrollment in Tennessee was much less than 30,000 students. To remove the discrepancy, 1986 district enrollment figures were reduced by 3 percent or 5 percent, depending on the difference reported between 1986 and 1987. A similar, but reversed, pattern was seen in 1986 Pennsylvania district enrollment figures, which were about 60,000 lower than school figures and 1987 figures. To remove this discrepancy, 1986 district enrollment figures were increased by 3 percent or 5 percent. Finally, 1991 district enrollments in Michigan were inconsistent with adjacent years and with school figures, so those figures were replaced with the corresponding school-level figures. Note that in every case in which district enrollment figures were modified, other figures that add to the enrollment (graded and ungraded enrollment and race/ethnicity counts) were modified proportionally.

14. After the corresponding imputation of teacher FTE data, described below, three additional global imputation steps were performed. First, all imputed district enrollment figures that were inconsistent with adjacent years (i.e., "V's") were re-imputed to be equal to school figures if that would remove the inconsistency. Second, remaining school enrollment discrepancies in districts in which teacher FTE counts were consistent were replaced using a constant student/teacher ratio in the district. Third, in all one-school districts with inconsistencies, the school enrollment was imputed equal to the district enrollment figure. At the completion of this step, about 55 discrepancies of more than 10 percent or missing school enrollments remain.

Matching the sum of FTE teacher counts in schools in a district to the FTE teacher total recorded on the district file. Many discrepancies in teacher enrollments, between school and district figures, are expected because some teachers are not assigned to particular schools and not counted in school staff counts. Of course, many teachers who split time between schools are counted in school staff FTE counts as well as district counts, but discrepancies are expected. . Therefore, at the conclusion of the editing and imputation, differences remain in the total enrollments based on school and district surveys (compare tables A4 and A14). Five steps were implemented for removing teacher FTE discrepancies.

2. Whenever a modification of a school's identification code was made to remove an enrollment discrepancy, that also tended to remove a corresponding teacher FTE discrepancy.

3. As a first systematic editing step, when student/teacher ratios were available for a school with a discrepancy in the FTE teacher count, based either in other years or at the district level in the same year, the teacher FTE count was imputed as the specified fraction of the school's enrollment.

4. Next, if student and teacher counts were available for other schools in the same district and year, a discrepancy was removed by applying that ratio to the students in the school.

5. Next, when it removed a discrepancy while not creating a new inconsistency with adjacent years at the district level, the district FTE count was replaced by the sum of school FTE counts.

6.   Finally, when no school-level student/teacher ratio was available, FTE teachers reported at the district level were allocated to schools in proportion to the schools' enrollments. At the conclusion of this step, 25 discrepancies remained.

7.   *Imputation of school race/ethnic counts.* This step made use of previously imputed district-level race/ethnic counts. (See appendix B, Round 3, Phase IV.) Editing and imputation were carried out on race percentages of membership and translated back into student counts and stored on SCHRACES.SD2 as a final step. First, all partially missing race counts were set to zero if races reported added up to more than 75 percent of the membership. Next, for all schools with race data, a mean percentage estimate for 1992-93 and an annual increase were estimated based on years with data. For all other schools, these statistics were imputed based on district (or if necessary, state) statistics, including an appropriate error percentage. Then percentages missing for individual years were imputed using the 1992-93 mean estimate and annual increase estimate, including an appropriate error percentage. Single-year discrepancies of more than 25 percent (and more than 50 students) were replaced with appropriate imputations, and 141 anomalies were individually edited.

No race data were reported (or imputed) for 1986-87; and districts in states with completely missing data had been imputed from a census data on percentages of school-aged children by race. The number of states not reporting race counts decreased from 17 in 1987-88 and 13 in 1988-89 to only one per year after 1992-93. Results for 1987-88 were evaluated by comparison with state counts available from the Office of Civil Rights for 1984-85 and 1986-87.

The numbers of imputed membership, FTE teacher values, and race/ethnic counts on the school file are shown in tables B10, B11, and B12.

Table B10.   Number of schools for which membership imputations were generated

	Imputed	Not Imputed
1986-87	714	83,535
1987-88	1,111	83,030
1988-89	784	83,255
1989-90	584	83,816
1990-91	386	84,953
1991-92	872	84,442
1992-93	1,259	84,127
1993-94	1,442	84,631
1994-95	1,209	85,570
1995-96	1,672	85,962
1996-97	3,944 [1]	84,766
1997-98	1,461	88,343
1998-99	1,467	89,825

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal School File. [1] Original CCD data are missing for New Jersey in 1996-97.

Table B11. Number of schools for which FTE teacher imputations were generated

	Imputed	Not Imputed
1986-87	7,152	77,097
1987-88	8,856	75,825
1988-89	13,654	70,385
1989-90	4,246	80,154
1990-91	6,925	78,414
1991-92	6,098	79,216
1992-93	7,520	77,866
1993-94	3,292	82,781
1994-95	5,095	81,684
1995-96	5,863	81,771
1996-97	9,482	79,228
1997-98	8,882	80,922
1998-99	4,193	87,099

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal School File.

Table B12. Number of schools for which race/ethnic imputations were generated

	Imputed	Not Imputed
1986-87	-	-
1987-88	21,917	62,224
1988-89	14,378	69,661
1989-90	10,614	73,786
1990-91	9,414	75,925
1991-92	7,285	78,029
1992-93	3,372	82,014
1993-94	753	85,320
1994-95	717	86,062
1995-96	785	86,849
1996-97	3,068	85,642
1997-98	794	89,010
1998-99	3,066	88,226

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal School File.

Editing and Imputation Steps for the Longitudinal District File

The longitudinal editing and imputation system for adding 1996-97 and 1997-98 data to the Longitudinal District File used procedures previously developed. Thus, the project may be considered Round 4 of the development of the Longitudinal District File. The editing and imputation in Round 3 was carried out primarily for the 1994-95 and 1995-96 school years, using procedures described in Appendix B. (The editing and imputation rules for the years preceding 1994-95 (Round 2) were developed earlier and are also summarized in Appendix B.)

The procedures for adding a year to district records consists of a series of five SAS programs, referred to in Appendix B as Phase I through Phase V. These programs were executed interactively, with examination of intermediate outputs followed by subsequent, more detailed, examination of small numbers of ambiguous cases. Missing values and values that were evaluated as very unlikely to reflect the actual status of education in school districts were replaced with statistically plausible values. The judgments to replace reported values were naturally very conservative, to avoid eliminating real variability in school district information.

The numbers of district values imputed for each year (cumulatively over four rounds of imputations) are given in tables B2, B3, and B4. The numbers of missing responses imputed are given in the first half of each table (B2a, B3a, B4a), and the number of non-missing values replaced are given in the second half of each table (B2b, B3b, B4b). Some of the replaced non-missing values are really indicators of missing data, however, so the division of imputation counts into the two types is somewhat arbitrary. For two examples, all missing data for many variables on the basic 1990-91 CCD district file were zero; and "00" was used to represent no reported (i.e., missing) grade span in several years.

Editing and Imputation Steps for the Longitudinal District Staff File

CCD began to collect staff breakdowns by category with the 1992-93 school year, and this effort has been a challenge because each state uses a unique categorization of staff that must be fit into the CCD schema. In many cases, categories did not fit and states did not report categories, even though one can be sure that staff in the categories were employed in schools and districts. Table B13 shows the states for which each category was reported either missing (M), not employed in the state (N), or all zeros (Z) in each year.

The method for imputing missing data was to estimate the linear trend for each district and to impute using the linear trend, adding in the appropriate error variance. The most difficult aspect of this imputation was the identification of "jumps," that is, years in which a district changed from reporting no staff in the category to reporting the staff category. Failing to eliminate these cases would distort the linear trend. On the other hand, it is reasonable to find that many districts change from zero to a positive number when they hire the first person in the particular category. The criterion for determining that a change from zero to a positive report was a reporting change was that the value following the last zero was at least three quarters of the value in ensuing year. Thus an increase from 0 to 10 to 20 was considered to be a real increase, while for an increase from 0 to 80 to 100, the zero was not considered in estimating the linear trend.

Table B13. States in which staff breakdowns are all missing or zero, by year

	Aides							Instructional Coordinators						
	92	93	94	95	96	97	98	92	93	94	95	96	97	98
Alabama			Z											
Alaska								M	M	M	M	M	M	M
Arizona														
Arkansas														
California			Z											
Colorado														
Connecticut														
Delaware														
District of Columbia	M					M		M					M	
Florida														
Georgia					Z	Z		M	M					
Hawaii														
Idaho														
Illinois					N	N	N							
Indiana														
Iowa														
Kansas														
Kentucky					Z							Z		
Louisiana								Z						
Maine														
Maryland														
Massachusetts														
Michigan														
Minnesota						M					M		M	
Mississippi														
Missouri														
Montana	M	M	M	M	M	M	M							
Nebraska														
Nevada														
New Hampshire								M	M	M	M	M	M	M
New Jersey					M							M		
New Mexico														
New York														
North Carolina														
North Dakota														
Ohio								M	M	M	M	Z	Z	
Oklahoma														
Oregon														
Pennsylvania														
Rhode Island														
South Carolina		M	M	M	N	N	N							
South Dakota														
Tennessee								M	M	M	M	N	N	N
Texas											M			
Utah														
Vermont	M	M						M	M					
Virginia	M				M	M	M					M	M	M
Washington	M							M	M	M	M	M	M	N
West Virginia														
Wisconsin														
Wyoming														
American Samoa														
Guam														
Northern Marianas														
Puerto Rico		M	M											
Virgin Islands								M	M					

Table B13. States in which staff breakdowns are all missing or zero, by year (cont.)

	Elementary Guidance Counselors							Secondary Guidance Counselors						
	92	93	94	95	96	97	98	92	93	94	95	96	97	98
Alabama														
Alaska														
Arizona														
Arkansas														
California														
Colorado														
Connecticut														
Delaware														
District of Columbia	M					M		M					M	
Florida														
Georgia	M	M						M	M					
Hawaii	M	M	M	N	N	N	M	M	M	M	N	N	N	M
Idaho														
Illinois														
Indiana														
Iowa														
Kansas														
Kentucky														
Louisiana				M	M						M	M		
Maine														
Maryland														
Massachusetts	M	M	M	M	M	M	Z	M	M	M	M	M	M	Z
Michigan	M	M	M	M	M	Z	Z	M	M	M	M	M	Z	Z
Minnesota				M		M					M		M	
Mississippi														
Missouri														
Montana														
Nebraska														
Nevada														
New Hampshire														
New Jersey					M							M		
New Mexico														
New York														
North Carolina	M	M	M	N	N	N	M	M	M	M	N	N	N	M
North Dakota														
Ohio														
Oklahoma														
Oregon														
Pennsylvania														
Rhode Island														
South Carolina														
South Dakota	M							M						
Tennessee														
Texas	M	M	M	M	M	M		M	M	M	M	M	M	
Utah														
Vermont	M	M						M	M					
Virginia					M	M	M					M	M	M
Washington	M							M						
West Virginia														
Wisconsin														
Wyoming														
American Samoa														
Guam														
Northern Marianas														
Puerto Rico														
Virgin Islands														

Table B13. States in which staff breakdowns are all missing or zero, by year (cont.)

	Total Guidance Counselors						
	92	93	94	95	96	97	98
Alabama							
Alaska							
Arizona							
Arkansas							
California							
Colorado							
Connecticut							
Delaware							
District of Columbia	M					M	
Florida							
Georgia							
Hawaii							
Idaho							
Illinois							
Indiana							
Iowa							
Kansas							
Kentucky							
Louisiana							
Maine							
Maryland							
Massachusetts							
Michigan	M						
Minnesota				M		M	
Mississippi							
Missouri							
Montana							
Nebraska							
Nevada							
New Hampshire							
New Jersey					M		
New Mexico							
New York							
North Carolina							
North Dakota							
Ohio							
Oklahoma							
Oregon							
Pennsylvania							
Rhode Island							
South Carolina							
South Dakota							
Tennessee							
Texas							
Utah							
Vermont	M	M					
Virginia					M	M	M
Washington	M						
West Virginia							
Wisconsin							
Wyoming							
American Samoa							
Guam							
Northern Marianas							
Puerto Rico							
Virgin Islands							

Table B13. States in which staff breakdowns are all missing or zero, by year (cont.)

	Library Specialists							Library Support						
	92	93	94	95	96	97	98	92	93	94	95	96	97	98
Alabama										Z				
Alaska														
Arizona														
Arkansas														M
California								M	M	Z	M	M	M	M
Colorado														
Connecticut														
Delaware														
District of Columbia						M		M				M		
Florida														
Georgia								M	M	M	M			
Hawaii														
Idaho														
Illinois												N	N	N
Indiana														
Iowa												Z		
Kansas														
Kentucky														
Louisiana														
Maine														
Maryland														
Massachusetts														
Michigan														
Minnesota				M		M						M		
Mississippi														
Missouri								M	M	M	M	N	N	N
Montana								M	M	M	M	M	M	M
Nebraska								M	M					
Nevada														
New Hampshire														
New Jersey				M								M		
New Mexico														
New York														
North Carolina								M	M	M	M	N	N	
North Dakota														
Ohio														
Oklahoma								M	M	M	M	M	M	N
Oregon														
Pennsylvania														
Rhode Island														
South Carolina									M	M	M	N	N	N
South Dakota									M	M	M			
Tennessee								M	M	M	M	N	N	N
Texas								M	M	M	M	M	M	M
Utah														
Vermont		M						M	M					
Virginia					M	M						M	M	M
Washington								M						
West Virginia										Z				
Wisconsin														
Wyoming														
American Samoa														
Guam														
Northern Marianas														
Puerto Rico														
Virgin Islands														

Table B13. States in which staff breakdowns are all missing or zero, by year (cont.)

	District Administrators							District Support Staff						
	92	93	94	95	96	97	98	92	93	94	95	96	97	98
Alabama										M				
Alaska														
Arizona														
Arkansas														
California										Z				
Colorado														
Connecticut														
Delaware														
District of Columbia	M					M		M					M	
Florida								M						
Georgia								M				Z	Z	
Hawaii														
Idaho														
Illinois				M				M	M	M	M	N	N	N
Indiana														
Iowa														
Kansas														
Kentucky					Z							Z		
Louisiana														
Maine								M	M	M	M	N	N	N
Maryland														
Massachusetts														
Michigan														
Minnesota	M	M	M	M	M	M							M	
Mississippi					M									
Missouri												M	M	
Montana								M	M	M	M	M	M	M
Nebraska														
Nevada														
New Hampshire		M	M					M	M	M	M	M	M	
New Jersey					M							M		
New Mexico														
New York														
North Carolina								M	M	M	M	N	N	
North Dakota														
Ohio														
Oklahoma														
Oregon														
Pennsylvania														
Rhode Island														
South Carolina									M	M	M	N	N	N
South Dakota														
Tennessee								M	M	M	M	N	N	N
Texas														
Utah														
Vermont	M	M						M	M					
Virginia					M	M		M				M	M	M
Washington	M							M						
West Virginia														
Wisconsin														
Wyoming														
American Samoa														
Guam														
Northern Marianas														
Puerto Rico														
Virgin Islands														

Table B13. States in which staff breakdowns are all missing or zero, by year (cont.)

	School Administrators							School Support Staff						
	92	93	94	95	96	97	98	92	93	94	95	96	97	98
Alabama			Z											
Alaska								M	M	M	M	M	M	
Arizona														
Arkansas														
California		Z												
Colorado														
Connecticut														
Delaware														
District of Columbia	M					M		M					M	
Florida														
Georgia					Z	Z		M	M					
Hawaii														
Idaho														
Illinois					N	N								M
Indiana														
Iowa														
Kansas														
Kentucky					Z							Z		
Louisiana								Z						
Maine														N
Maryland														
Massachusetts														
Michigan														
Minnesota						M					M		M	
Mississippi														
Missouri														
Montana	M	M	M	M	M	M								
Nebraska														N
Nevada														M
New Hampshire								M	M	M	M	M	M	
New Jersey					M							M		
New Mexico														
New York														
North Carolina														
North Dakota														
Ohio								M	M	M	M	Z	Z	
Oklahoma														
Oregon														
Pennsylvania														
Rhode Island														
South Carolina		M	M	M	N	N								N
South Dakota														
Tennessee								M	M	M	M	N	N	
Texas										M				
Utah														
Vermont	M	M						M	M					
Virginia	M				M	M	M					M	M	M
Washington	M							M	M	M	M	M	M	
West Virginia														
Wisconsin														
Wyoming														
American Samoa														
Guam														
Northern Marianas														
Puerto Rico		M	M											
Virgin Islands									M	M				

Table B13. States in which staff breakdowns are all missing or zero, by year (cont.)

	Student Support Staff							Other Support Staff						
	92	93	94	95	96	97	98	92	93	94	95	96	97	98
Alabama										Z				
Alaska	M	M	M	M	M	M								
Arizona														
Arkansas														
California										Z				
Colorado														
Connecticut														
Delaware														
District of Columbia	M					M		M					M	
Florida	M							M						
Georgia	M	M						M						
Hawaii														
Idaho														
Illinois												N	N	M
Indiana														
Iowa														
Kansas														
Kentucky					Z							Z		
Louisiana														
Maine														
Maryland														
Massachusetts														
Michigan														
Minnesota						M							M	
Mississippi														
Missouri													M	M
Montana	M							M	M	M	M	M	M	M
Nebraska														
Nevada														
New Hampshire	M	M	M	M	M	M	M							
New Jersey				M								M		
New Mexico														
New York														
North Carolina														
North Dakota														
Ohio														
Oklahoma	Z	M	M	Z	Z	Z								
Oregon														
Pennsylvania														
Rhode Island														
South Carolina	M	M	M	M	N	N	N		M	M	M	N	N	N
South Dakota														
Tennessee	M	M	M	M	N	N	N							
Texas														
Utah														
Vermont	M	M						M	M					
Virginia	M				M	M	M	M				M	M	M
Washington	M	M	M	M	M	M	N	M						
West Virginia														
Wisconsin														
Wyoming														
American Samoa														
Guam														
Northern Marianas														
Puerto Rico														
Virgin Islands														

Note: Source: U.S. Dept. of Education, National Center for Education Statistics, Common Core of Data, Longitudinal District Staffing File. M indicates that all reported values are missing in the state. Z indicates all zeros.

Appendix C. Contents of the Longitudinal CCD District SAS Files

The files are SAS files. There are 29 files: 13 individual-year local agency-level files, labeled AIRLEA86 through AIRLEA98; 13 individual-year school-level files, labeled AIRSCH86 through AIRSCH98; 1 thirteen-year combined agency-level file, LEA13YRS, 1 thirteen-year combined school-level file, SCH13YRS, and 1 agency-level staffing category file with seven years' data. The contents of AIRLEA98 and AIRSCH98 are given below, as examples of the individual-year files. The combined-year files have variables of the same names.[7]

The names of all variables that change values each year (e.g., MEMBER – enrollment) include a two digit name of the year. For example, on the 1998 file and the combined year file, the variable MEMBER98 appears. For the district data, LEAID does not change over years, and for the school data, MASTERID does not change over years. On the combined year school file, schools that change districts are represented by multiple records with the same MASTERID, but different values for LEAID. The variable LYRS indicates which years the school is in each LEA.

In 1998-99, three variables in the database for the period 1987-88 through 1997-98, REGDIPLO, OTHDIPLO, and OTHHSC (regular and other diploma recipients and other high school completers) were removed from the file and replaced by AMDPL, AMOHC, ASDPL, ASOHC, BLDPL, BLOHC, HIDPL, HIOHC, WHDPL, WHOHC, MALDPL, MALOHC, FEMDPL, FEMOHC, TOTDPL, and TOTOHC (race and gender breakdowns and totals for diploma recipients and other high school completers). The single variable, FLE, was replaced by three variables, FLE, REDLCH, and TOTFRL (free lunch, reduced price lunch, and combined eligible student counts). Counts of LEP and MIGRANT students were added, although migrant counts were not edited or imputed in this version of the Longitudinal Database. Indicators of charter, magnet, Title I, and schoolwide Title I schools were added to CCD in 1998-99, and they are presented without further editing or imputation in the Longitudinal Database. Finally, CCD directory information was enhanced in 1998-99 by inclusion of both location and mailing addresses for schools and districts.

---

[7] One exception is that AM on the school file and AMIND on the district file both refer to counts of American Indians and Alaska Natives.

Contents of 1998 School File Component of Longitudinal Database (SAS Format)

Variable	Type	Length	Notes	Label
AM98	Number	8	1	American Indian/Alaskan Native Students
ASIAN98	Number	8	1	Asian/Pacific Islander Students
BLACK98	Number	8	1	Black Non-Hispanic Students
CHARTR98	Char	1	2	Charter School
FIPST	Char	2		Fips State Number
FRELCH98	Number	8	1	Free Lunch Eligible Students
FRLEM98	Char	1	2	Imputation flag for FLE, REDLCH, TOTFRL
FTE98	Number	8	1	Classroom Teachers (Full-Time Equiv)
FTEM98	Char	1		Imputation flag Teachers (FTE)
GSHI98	Char	2		School's highest grade (from gr enrl)
GSLO98	Char	2		School's lowest grade (from gr enrl)
HISP98	Number	8		Hispanic Students
LCITY98	Char	30		Location City Name
LEAID	Char	7		NCES School District Code
LEANM98	Char	60		Name of Operating Agency
LOCALE98	Char	1		Locale Code
LSTATE98	Char	2		Location USPS State Abbreviation
LSTREE98	Char	30		Location Address
LYRS	Char	13	3	Y) school in LEA, N) not
LZIP98	Char	5		Location 5-Digit Zip Code
LZIP498	Char	4		Location Zip+4 (if assigned)
MAGNET98	Char	1	2	Magnet School
MASTERID	Char	12	3	Permanent School Identification Number
MCITY98	Char	30		Mailing City Name
MEMBER98	Number	8	1	Students (Total Reported Membership)
MEMBM98	Char	1	3	Imputation flag Student Membership
MIGRNT98	Number	8	2	Migrant Students
MSTATE98	Char	2		Mailing USPS State Abbreviation
MSTREE98	Char	30		Mailing Address
MZIP98	Char	5		Mailing 5-Digit Zip Code
MZIP498	Char	4		Mailing Zip+4 if Assigned
NCESSCH	Char	12		NCES School Code in year 1997-98,
PHONE98	Char	10		Telephone Number of School
RACEM98	Char	1	3	Imputation flag Race/ethnicity
REDLCH98	Number	8	12	Reduced-price Lunch El Students
SCHNAM98	Char	50		School Name
SEASCH98	Char	20		State School ID
STATUS98	Char	1		Operational Status Code
STID98	Char	14		State Agency ID
STITLI98	Char	1	2	School-wide Title 1
SYRS	Char	13	3	Y) school open, N) closed
TITLEI98	Char	1	2	Title 1 Eligible School
TOTFRL98	Number	8	12	Total Free and Reduced Lunch Students
TYPE98	Char	1		School Type Code
WHITE98	Number	8	1	White Non-Hispanic Students
YR	Number	8		School year (Fall)

Notes: (1) Edited number. (2) New in 1998-99. (3) Variable created for the longitudinal file.

Contents of 1998 District (LEA) File Component of Longitudinal Database (SAS Format)

Variable	Type	Length	Notes	Label
AMDPL98	Number	8	12	Alaskan/Amer Ind Diploma Recipients
AMIND98	Number	8	4	Alaskan/American Indian Students
AMOHC98	Number	8	12	Alaskan/Amer Ind Other HS Completrs
ASDPL98	Number	8	12	Asian Diploma Recipients
ASIAN98	Number	8	4	Asian Students
ASOHC98	Number	8	12	Asian Other HS Completers
BLACK98	Number	8	4	Black Students
BLDPL98	Number	8	12	Black Diploma Recipients
BLOHC98	Number	8	12	Black Other HS Completers
BOUND98	Char	1		Operational Status Code
CMSA98	Char	6		CMSA/PMSA/MSA Code
CONAME98	Char	30		County Name
CONUM98	Char	5		Fips County Number (Fipst+County)
FEMDPL98	Number	8	12	Female Diploma Recipients
FEMOHC98	Number	8	12	Female Other HS Completers
FIPST	Char	2		Fips State Number
FLE98	Number	8	4	Free Lunch Eligible Students
FLEM98	Char	1	3	Imputation flag: Free Lunch Eligible Students
FRLEM98	Char	1	3	Imputation flag: Free & Red Price Eligibles
FTE98	Number	8	1	Total Fte Teachers
FTEM98	Char	1	3	Imputation flag: FTE teacher count
GSHI98	Char	2	1	Agency High Grade Offered
GSHIM98	Char	1	3	Imputation flag: High Grade
GSLO98	Char	2	1	Agency Low Grade Offered
GSLOM98	Char	1	3	Imputation flag: Low Grade
HIDPL98	Number	8	12	Hispanic Diploma Recipients
HIOHC98	Number	8	12	Hispanic Other HS Completers
HISP98	Number	8	4	Hispanic Students
LCITY98	Char	30		Location City Name
LEAID	Char	7		NCES School District Code
LEP98	Number	8	1	Limited-English-Proficient Students
LEPM98	Char	1	3	Imputation flag: Limited English Proficient
LOCALE98	Number	5	4	Locale: Weighted. Maximum of School Locales
LOCALM98	Char	1	3	Imputation flag: Locale
LSTATE98	Char	2		Location USPS State Abbreviation
LSTREE98	Char	30		Location Address
LZIP98	Char	5		Location 5-Digit Zip Code
LZIP498	Char	4		Location Zip+4 (if assigned)
MALDPL98	Number	8	12	Male Diploma Recipients
MALOHC98	Number	8	12	Male Other HS Completers
MCITY98	Char	30		Mailing City Name
MEMBEM98	Char	1	3	Imputation flag: Enrollment
MEMBER98	Number	8	1	Total Calculated Students
MFDPLM98	Char	1	3	Imputation flag: Gender of Diploma Recipients
MFOHCM98	Char	1	3	Imp. flg Gender of Other HS Completers **(continued)**

Contents of 1998 District (LEA) File Component of Longitudinal Database (SAS Fmt) continued

Variable	Type	Length	Notes	Label
MSC98	Char	1		Metro Status Code
MSTATE98	Char	2		Mailing USPS State Abbreviation
MSTREE98	Char	30		Mailing Address
MZIP98	Char	5		Mailing 5-Digit Zip Code
MZIP498	Char	4		Mailing Zip+4 (if assigned)
NAME98	Char	60		Name Of Local Education Agency
NXTYRID	Char	7	3	For closing LEAs: ID of successor
PHONE98	Char	10		Telephone Number Of Agency
PK1298	Number	8	1	Total PK thru 12 Students
PK12M98	Char	1	3	impflg Students in Grades PK to 12
PPOV90	Number	8	1	Pct chldrn in poverty (Census 90)
PPOV95	Number	8	1	Pct chldrn in poverty (Census 95)
PPOV90M	Char	1	3	impflg Pct chldrn pov (Census 90)
PPOV95M	Char	1	3	impflg Pct chldrn pov (Census 95)
PRVYRID	Char	7	3	ID of closed LEA that sent students
RACDPM98	Char	1	3	impflg Race/Ethnic Diploma Recips
RACEM98	Char	1	3	impflg Race/Ethnic counts
RACOCM98	Char	1		impflg Race/Ethnic Other HS Complet
REDLCH98	Number	8	12	Reduced Price Lunch Eligibles
SCH98	Number	8	1	Number Of Schools (School Univ)
SCHM98	Char	1	3	impflg Number of Schools
SPECED98	Number	8	1	Count Of Special Ed IEP Students
SPECEM98	Char	1	3	impflg Special Education Students
STID98	Char	14		State Agency ID
SYFALL	Char	2		School Year (Fall)
TOTDPL98	Number	8	12	Diploma Recip (Total)
TOTDPM98	Char	1	3	impflg Total Diploma Recipients
TOTFRL98	Number	8	12	Free &d Red Price Lunch Eligibles
TOTOHC98	Number	8	12	Other HS Completers (Total)
TOTOHM98	Char	1	3	impflg Total Other HS Completers
TYPE98	Char	1		Agency Type Code
UG98	Number	8	1	Total Ungraded Students
UGM98	Char	1	3	impflg Ungraded Students
UNION98	Char	3		Supervisory Union Number
WHDPL98	Number	8	12	White Diploma Recipients
WHITE98	Number	8	4	White Students
WHOHC98	Number	8	12	White Other HS Completers
YEAR	Number	8		Year CCD collected (Fall of Sch Yr)
YRS	Char	13	3	Y)ears present, M)ssng, N)o schls

Notes: (1) Edited number. (2) New variable in 1998-99. (3) Variable created for the longitudinal file. (4) Aggregated from edited school file.

Contents of 1998 School File Component of Longitudinal Database (SAS Format)

Variable	Type	Length	Notes	Label
AIDES	Number	8		Instructional Aides
CORSUP	Number	8		Instructional Coordinators/Supervisors
EGM	Char	1		Flag: Elementary Guidance Counselors
ELMGUI	Number	8		Elementary Guidance Counselors
ELMTCH	Number	8		Elementary Teachers
KGTCH	Number	8		Kindergarten Teachers
L2M	Char	1	1	Imputation flag: Library Media Support Staff
LAM	Char	1	1	Imputation flag: LEA Administrators
LEAADM	Number	8		LEA Administrators
LEAID	Char	7		Unique System ID (NCES Assigned)
LEASUP	Number	8		LEA Administrators Support Staff
LEM	Char	1	1	Imputation flag: LEA Support Staff
LIBSPE	Number	8		Librarians/Media Specialists
LIBSUP	Number	8		Library Media Support Staff
LSM	Char	1	1	Imputation flag: Librarians/Media Specialists
OSM	Char	1	1	Imputation flag: All Other Support Staff
OTHSUP	Number	8		All Other Support Staff
PKTCH	Number	8		Prekindergarten Teachers
SAM	Char	1	1	Imputation flag: School Administrators
SCHADM	Number	8		School Administrators
SCHSUP	Number	8		School Administrators Support Staff
SECGUI	Number	8		Secondary Guidance Counselors
SECTCH	Number	8		Secondary Teachers
SEM	Char	1	1	Imputation flag: School Admin Support Staff
SGM	Char	1	1	Imputation flag: Secondary Guidance Counselors
SSM	Char	1	1	Imputation flag: Student Support Services Staff
ST	Char	2		State USPS Code
STFYRS	Char	7	1	Years LEA on file (Y), or not (N): 92-98
STUSUP	Number	8		Student Support Services Staff
TAM	Char	1	1	Imputation flag: Instructional Aides
TGM	Char	1	1	Imputation flag: Total Guidance Counselors
TIM	Char	1	1	Imputation flag: Instructional Coord./Supervisors
TM	Char	1	1	Imputation flag: Total FTE Teachers [LEA]
TOTGUI	Number	8		Total Guidance Counselors
TOTTCH	Number	8		Total FTE Teachers [LEA]
TTM	Char	1	1	Imputation flag: FTE Teacher by level
UGTCH	Number	8		Teachers of Ungraded Classes

Notes: (1) Variable created for the longitudinal file.

References

Levine, R.E., McLaughlin, D.H., & Sietsema, J. (1996). *Trends in school district demographics, 1986-87 to 1990-91.* Washington DC: U.S. Department of Education, National Center for Education Statistics: NCES 96-399.

McLaughlin, D.H. (1999). A Ten-Year Longitudinal CCD Local Education Agency Non-Fiscal Survey File. 1986/87 – 1995/96. PDF available from author.

McLaughlin, D.H., Huberman, M.B., Hawkins, E.K. & Hoffman, L.M. (1997). *Characteristics of small and rural school districts.* Washington DC: U.S. Department of Education, National Center for Education Statistics: NCES 97-529.

McLaughlin, D.H. (2001). A Twelve-Year Longitudinal CCD Non-Fiscal Survey Database: 1986/87 - 1997/98. PDF at http://nces.ed.gov/ccd/pau12yr.html.