

CHANGES IN THE USE OF EDUCATION ADMINISTRATIVE RECORDS: THE COMMON CORE OF DATA

Lee M. Hoffman and Beth Aronstamm Young
National Center for Education Statistics, U.S. Department of Education

Abstract

This paper will discuss the evolution of Administrative Records collections from simple lists to a major, multi-purpose source of data to meet increasing needs from both the statistical and non-statistical communities. We will argue that changes in the data collected by administrative records, the users of the data, and the products demanded by the public constitute systemic changes in how administrative records are seen and used. We will use the Common Core of Data (CCD), the annual NCES administrative records collection of data from every public school, district and state, to illustrate these changes.

Keywords: Administrative records, Statistical uses

The survey collection that today comprises the Common Core of Data (CCD) has been in existence for more than 35 years, and actually predates the agency that now houses it, the National Center for Education Statistics. The purpose of this paper is to examine the content and uses of the CCD surveys, in particular the Nonfiscal components, as these have changed in the last decade and to argue that the role of administrative records data, as exemplified by the CCD, is undergoing a substantial change.

Description of the CCD Surveys

The CCD collects a limited number of statistics about the universe of public elementary and secondary education institutions. The present system consists of four nonfiscal and two fiscal surveys. The nonfiscal surveys include the State Nonfiscal survey; the Public Education Agency (school district) Survey; the Public School survey; and the Early Estimates survey. (This last survey includes two items on school finance.) The fiscal surveys include the state-level National Public Education Finance Survey, and the Annual Survey of Local Governments School Systems (Form F-33). The F-33 belongs properly to the Bureau of the Census, but NCES has co-sponsored the collection, incorporating some of its own items and fiscal conventions, since 1990.

Collection Processes

For all surveys, regardless of the level at which information is reported, the state education agency (SEA) is the respondent. Participation is voluntary and SEAs receive no compensation for their work. Respondents include the 50 states, the District of Columbia, the Bureau of Indian Affairs, both overseas and domestic Department of Defense Dependents Schools, and the five outlying areas of American Samoa, the Commonwealth of the Northern Marianas Islands, Guam, Puerto Rico, and the U.S. Virgin Islands. With the

exception of BIA and DoDDS, which do not report fiscal data, there is item nonresponse but no survey nonresponse.

The school and school district surveys, which are collected as flat files with the option of using editing software, are distributed in the form of prior year files in late January of the school year reported. Data collection is cut off almost nine months after this mail out on September 30; and the ongoing processing, editing, respondent review, and correction continue for another three months. The SEA reports most items that appear on the final CCD files. A few, such as the unique identification number for each school and agency, are assigned by NCES.¹

Upon first receipt, the files are checked for valid cases that have sufficient information for further processing. Identification numbers are assigned to new schools and agencies. Data are not edited at this point. An example of data processing at this stage would be to determine whether a school has an acceptable address and operational status, but not to examine the numbers of teachers and students reported for the school.

Data are edited during the next editing stage, after a file has been determined to contain only valid cases. As a final step, items not reported by the SEAs, such as locale code or county FIPS code, are added. Missing data are rarely imputed, and challenged data, if asserted to be correct by the SEA, are suppressed only if they are impossible. An example of “impossible data” would be a state’s report that it has no students who were granted General Education Development (GED) diplomas. In this case, editing would change the reports of “not applicable” to “missing,” because external evidence shows that all states grant such degrees.

In the 1999-2000 school year, the CCD reported data for more than 92,000 public schools and almost 17,000 local education agencies.

Content of the CCD Surveys

The nonfiscal CCD surveys provide three types of information:

- Directory information, such as school or agency name, address, county, telephone number;
- Descriptive characteristics of the school or agency such as its type (example, regular school district or one providing special administrative services), urbanicity (example, in a rural or urban locale), and program “flags” (example, magnet, charter, Title I school); and
- Statistics about the students and staff associated with the school or agency (example, numbers of students by racial/ethnic group by grade).

The survey content is nested to some degree. Each agency must be associated with a state, and each school must have an associated agency. However, data are not expected to necessarily sum from one level to the next. For example, the number of students reported

¹ The Bureau of the Census, Governments Division, is the collection and processing agent for the CCD. References to NCES’s role in the CCD include the survey activities of Census.

for a local education agency may differ from the sum of students reported for its schools because there may be some students served outside a school setting or because the schools may report students tuitioned in from some other district.

Changes in the CCD – Late 1980's to Present

This paper focuses on changes made during the 1990's, subsequent to external review of NCES and the CCD survey system. Acting on recommendations made in the mid-1980's, NCES had by 1989 revised the CCD survey system.² Since that time the content of the CCD has expanded and NCES has worked with its respondent and user groups to introduce data definitions and associated standards that go considerably beyond the scope of the CCD surveys. The uses of CCD data have been affected primarily by changes in three areas: the content of the surveys; access to the information; and groups involved in survey development.

Changes in Survey Content

The CCD school and agency surveys of 1990 were relatively short. The School Universe Survey reported school identification numbers, mailing address, and telephone number; identified the agency with which the school was associated, school type (regular, vocational, etc.), locale type (rural, urban, etc.) and whether the school was new, closed, or ongoing. It also reported the enrollment by grade, total enrollment by race, total teacher FTE, and number of students eligible for free lunch.

The Local Education Agency Survey provided similar identification and directory information as well as agency type (regular school district, etc.), and several geographic indicators such as the FIPS county code, CMSA/PMSA/MSA status, and Metropolitan Status Code. Boundary change codes indicated whether agencies were new, closed, or reconfigured, and the survey included total ungraded and Prekindergarten (PK) through grade 12 enrollment, numbers of students with special education Individualized Education Programs (IEPs), and numbers of high school completers.

In the 1991-92 school year the School Universe Survey was unchanged but the Local Education Agency Universe Survey expanded considerably by adding two statistics. These were staff full-time equivalent (FTE) counts in a total of 17 instructional, support and administrative categories and counts of dropouts, by sex and race/ethnicity, in each of grades 7 through 12.

The staff data items had been collected in previous years on the State Nonfiscal Survey (and remain on that survey). The dropout statistic was new. Both statistics, particularly dropouts, required considerable negotiation between NCES and SEAs to achieve acceptable comparability among state reports.

² Internal discussion draft, February 1991, response to Hall, G., Jaeger, R., Kearney, P. and Wiley, D, *Alternatives for a National Data System on Elementary and Secondary Education*, National Academy of Sciences, December 1985.

The content of the CCD surveys remained unchanged until the 1998-99 school year. At that time several items were added to the Local Education Agency and School Universe Surveys. The changes were intended to make the surveys more useful as sampling frames for research and program evaluation purposes. For both surveys, this involved adding a location address (in addition to mailing address) and asking for the grade span (which had previously been determined by actual student enrollment). The agency survey added student detail: limited English proficient and summer migrant students served, and race and sex of high school completers. The school survey added “flags” identifying Title I, Title I school wide program, magnet, and charter schools. Student data included race and sex detail on grade-level enrollments, separate reporting of the number of free lunch and reduced-price lunch eligible students, and cumulative migrant student enrollment during the school year.

Changes in Data Access

Changing technology and the advent of the World Wide Web drastically altered the amount of access data users had to CCD information. Technology allowed NCES to publish its data files in user-friendly formats on the Internet. Several data tools, such as the CCD School and District Locator, gave the public access to information about individual schools and school districts. The Web opened the door to NCES data to parents, students, and the general public.

Paradoxically, the act of making CCD data more accessible to non-statisticians contributed to data quality improvements that made the files more useful to those with statistical and research interests. Increased access increased the CCD user group exponentially, and they in turn contributed questions, challenges, and suggestions on all aspects of the CCD survey system. This has led to better data quality and improvements in the editing process. When federal programs use the CCD for purposes such as determining grant eligibility this also increases the pressure on SEAs to report all data items.

Groups Involved in Survey Development

The changes that appeared in the 1991 surveys were very much concerned with the capacity to collect standardized and complete information from SEAs. The Council of Chief State School Officers, under contract with NCES, examined each data item reported by SEAs, identifying gaps and discrepancies and developing individualized state plans for CCD reporting improvement. NCES worked directly with SEAs to develop a consensus definition of school dropout that could be collected through the CCD. The dropout statistic was an innovation in two ways: it deliberately allowed state and district comparisons on a high-stakes educational outcome, and it very visibly required states to change their own definitions and reporting practices in order to comply with the CCD.³

In contrast, the 1998 survey expansions were motivated primarily by requests from those who used the surveys as sampling frames. This included statistical collections, such as the National Assessment of Education Progress, that ran into problems when a school was not

³ This paper does not discuss the NPEFS. However, it should be noted that during this same time NCES undertook considerable effort to ensure that state finance reports followed common standards.

located at its mailing address, and Department of Education program offices that wanted sub-state detail on program participation. This set of changes can be characterized as not requiring SEAs to change the content of their information systems (as had been the case with the dropout statistic) but requiring them, in many cases, to change the way in which these systems were organized. For example, the data needed for the program participation and school flag items are rarely maintained in a single data system within an SEA.

Data Standards Movement

At the same time that CCD content and use were expanding, the education information community was moving toward an endorsement of data sharing and national data standards. By 1989, NCES had established a National Cooperative Education Statistics System that included all state education agencies and those federal agencies and national associations that were major providers or consumers of PK-12 education data.

Under the Cooperative System, promulgating data standards has grown into an ongoing function for NCES. The agency has revised its Student and Staff Data Handbooks that had been out of print since the 1970's, and has established a system for updating them annually. The handbook, Financial Accounting for Local and State School Systems has been updated periodically, and is now in the process of complete revision.

These handbooks are much broader in scope than the content of the CCD surveys. Members of the Cooperative System are also creating new handbooks. At this time, work is underway on defining data elements in the areas of school crime, violence and disciplinary incidents; school facilities; and technology.

In addition to defining data elements, the Cooperative System has developed, or is influencing, electronic data standards. The group collaborated with postsecondary education to produce standards for transmitting electronic data about students and staff, and is participating with education software vendors in the development of extensible markup language (XML) standards as well. Finally, the Cooperative System has developed and published what are characterized as "best practice" guidelines for structuring state and local education data systems.

Current Uses of the CCD

CCD data have become increasingly available through electronic products in the last five years. The increased accessibility, in combination with work on standardizing data, has increased the ways in which the surveys are used. Four major categories of use include:

- Listing schools and agencies;
- Providing a sampling frame;
- Presenting basic descriptive data;
- Telling a story.

Listing Schools and Agencies

This is a longstanding function of the CCD. Commercial mailing list providers, such as MDR or QED, use it as the starting point for their own databases. Vendors rely on the CCD for mailing addresses. Police departments have called upon the CCD staff to track down the school that issued a class ring found at the scene of a crime, and film companies routinely check to ensure that they don't inadvertently use a real high school's name in a movie. With the introduction of easily customized mailing lists on the NCES Web Site, noncommercial use has increased.

Sampling Frames

The quality of the CCD data has improved in the last 10 years, and NCES now makes preliminary files available for sampling purposes under controlled conditions. These changes, and the content added in 1998, have increased the surveys' value as sampling frames. Within NCES, major surveys such as the National Assessment of Education Progress, the Early Childhood Longitudinal Study, and the Schools and Staffing Survey now use the CCD as a sampling frame.

Basic Data

CCD data are used extensively in NCES's flagship publications such as the *Digest of Education Statistics* and in multiple annual CCD data releases and less frequently published *State Profiles*. News media and education program planners use the CCD's basic data extensively, usually in combination with other information. CCD files provide information over time on school enrollments, numbers of teachers, teacher/pupil ratios, revenues and expenditures for public education, and similar bread-and-butter statistics. One tool created in the 1990's for non-technical audiences is the Public School and District Locator. This is an easy to use Web tool for finding information about individual institutions. Next to NCES's home page, and the Department of Education's postsecondary student financial aid site, the Locator is the most frequently visited Web site for the Department of Education.

Telling a Story

Additional content has meant additional interest. In recent years CCD data have supported reports on topical issues. These have included, for example, an analysis of the characteristics of small and rural schools and districts, a study of changes in racial isolation (the chances that a minority student will have non-minority classmates) within schools over time, and disparities in school district spending. There are also annual NCES reports of a more analytical nature that rely on the CCD. Examples include the annual *100 Largest School Districts*, and *Dropout Rates in the United States*, as well as an upcoming report on high school completion and dropout rates from 1991 to the present.

Emerging Uses of the CCD

There is a healthy interaction between the CCD and its products: better quality, more accessible data lead to survey improvements that in turn result in better, more accessible data. The survey system also benefits from having its respondents (SEAs) play a major role in redesign.

Education Institution Verification

Some parts of the U.S. Department of Education have begun to use the CCD in establishing program eligibility. Legislation for the Rural Education Achievement Program (REAP) limits participation to districts whose schools are classified as rural by the CCD. The Telecommunications Act of 1996 program (commonly known as E-Rate) requires that grantees include their CCD school identification code on applications. These uses have brought the CCD's data and coverage under closer scrutiny and arguably improved the survey products. The movement has been toward more use of the CCD identification codes to validate an institution's legitimacy by the Department of Education, and even some privately funded grants. The CCD identification codes also provide a means through which other data collections can link their files with the CCD and thereby increase the information in them.

On-Going Data Standards and Technology Change

The Office of the Chief Information Officer is considering using CCD definitions as standards for program data collections throughout the U.S. Department of Education. At the same time, program offices are attempting to coordinate their collections in order to reduce the reporting placed on schools and districts and to expand the potential uses of their own data. These offices are using CCD definitions wherever applicable, and will adopt the CCD school and agency identification codes. This will allow them to connect their data to the CCD, providing more information without additional cost and burden. Current legislation specifies the CCD locale codes for determining rural status under one program, and at least one draft of federal education legislation would incorporate the CCD dropout definition.

As mentioned earlier, the CCD data items are among those NCES standards that have been, or are being, endorsed by national groups. The American National Standards Institute X-12 Committee has approved the NCES-sponsored standards for transmitting information electronically about students and staff. NCES through its Cooperative System members is working to ensure that NCES/CCD data definitions are incorporated in the Schools Interoperability Framework that will support commercial school information management systems in areas ranging from students information to food service management.

CCD as a Research Data Base

"Research" is used loosely here to include almost any use of information to tell a story. In the past CCD data carried limited information. They said little about the health of the education system, or how it had changed, and their descriptive power was almost one-

dimensional. Having said a school has 23 teachers and 315 students, one has not said very much.

Developments within the CCD and in the broader world of information technology have gone a long way to change that situation. The SEAs that report CCD data are themselves maintaining information at a greater level of granularity. Almost all states maintain school-level report cards on their Web sites. Increasing numbers of states are instituting student record systems. Reducing the level of aggregation to the school – and at some time, student – level makes it possible to cut the data in a variety of ways with relatively little effort. Information technology also encourages the adoption of standards. To the extent that common definitions are applied to school-level data, the possibility of using these data to answer questions comes closer to being a reality.

Within the last year NCES has completed a longitudinal file (1986 – 1997) of CCD school and agency data. Unlike the official year-by-year files, this database links schools and agencies over time and imputes missing data. NCES plans to impute and add data for subsequent years to this file.

The CCD has two potential advantages as a research tool. One is simply that it provides a unique identifier for every public school and local education agency in the United States. The other advantage is that, as a voluntary administrative records collection, it is likely to include data items that are feasible to report and that are potentially valid and reliable. Consensus definitions must take into account and bridge differences in local practice.

What Is Needed to Increase CCD Use

The CCD has developed considerably over the past 10 years, but it is by no means ready to answer all the information needs of educational programs. Discussions with its users within SEAs and the Department of Education have identified at least four areas that need improvement.

1. Fill in all the holes. States must achieve 100 percent item response, or missing data must be imputed. If the latter occurs, NCES will probably have to maintain – and distinguish between – an imputed research file and a public use file that has some missing data.
2. Level the playing field. Administrative records data are of limited use unless they can assure users that all respondents follow the same definitions and reporting standards. This can be accomplished to some degree through “crosswalks” that equate data after the fact, but that is not as satisfactory a solution as common adoption of standards by all states.
3. Differentiate the cases. Users have requested additional items flagging various types of schools and districts, such as identifying various levels of school involvement in vocational education, to allow them to draw more specialized samples.
4. Add detail to existing statistics. There have been requests for information about the race/ethnicity and gender of education staff and for the ability to relate expenditure data to basic program areas, such as the regular curriculum or special education.

Strengths and Liabilities of Administrative Records Data

Do administrative records have sufficient potential as sources of valid and useful data to warrant the cost (both in federal dollars and respondent burden) of improving them? We would argue that they do, and that the wider adoption of data standards and the development of tools that make information increasingly accessible, support this argument. Administrative record collections such as the CCD have some built-in limitations. The derivative federal report will always have a lower priority than local or state information needs. Comparability will be limited by both human interpretation and by the variation in state education systems – there is no “tidy” way to deal with the geography of county school districts versus those that may be scattered over, but not coterminous with, several counties, or that may occupy the same space as other school districts. Access to universe data also raises the risk of divulging individual identity. One reason that the CCD does not include email addresses is that they may have a name embedded in them, and would thus breach NCES’s confidentiality standards.

However, administrative records provide data at little extra cost to the collector or burden on the respondent. They can serve as a lynchpin for connecting other data collections, thereby increasing information power considerably, again at little cost. Administrative records data derive from information used in the daily operation of schools, and are thus familiar to the public. The fact that they are used initially for practical education decisions suggests that they are likely to be scrutinized closely, which argues for better quality.

Content of Common Core of Data School Universe Survey, 1990 - Present

Pre-1990	1991	1998
State Agency ID	Same	Same
NCES Agency ID	Same	Same
Agency Name	Same	Same
School Name	Same	Same
State School ID	Same	Same
NCES School ID	Same	Same
Mailing Address, City, State, ZIP	Same	Same
		Location Address, City, State, ZIP
Telephone	Same	Same
School Type	Same	Same
Operational Status Code	Same	Same
		Grade Span (lowest, highest)
Locale Type Code	Same	Same
		Title I School Flag
		Title I Schoolwide Flag
		Magnet School Flag
		Charter School Flag
Teacher FTE	Same	Same
Enrollment by Grade (15 items)	Same	Enrollment by grade by race by sex
Total Free Lunch Eligible	Same	Same
		Total Reduced-Price Meal Eligible
Total Enrollment by Race	Same	See, Enrollment
		Migrant Students School Year

Content of Common Core of Data Local Education Agency Universe Survey, 1990 - Present

Pre-1991	1991	1998
NCES Agency ID	Same	Same
State Agency ID	Same	Same
Agency Name	Same	Same
Mailing Address, City, State, ZIP	Same	Same
		Location Address, City, State, ZIP
Telephone	Same	Same
Agency Type	Same	Same
Supervisory Union ID	Same	Same
FIPS County Code	Same	Same
		County Name
CMSA/PMSA/MSA	Same	Same
Metro Status Code	Same	Same
Boundary Change Code	Same	Same
		Grade Span (lowest-highest)
Ungraded Students	Same	Same
Total PK-12 Students	Same	Same
IEP Students	Same	Same
		Total Limited English Proficient Students
		Total Migrant Students Summer Programs
Regular Diploma Recipients	Same	By Race and Sex
Other Diploma Recipients	Same	By Race and Sex
Equivalency Recipients	Deleted – state level only	Same
Other Completers	Same	By Race and Sex
	Dropouts by Race by Sex by Grade 7-12	Same
	Teachers (6 items)	Same
	Aides	Same
	Instructional Coordinators	Same
	Counselors (3 items)	Same
	Library/Media Specialists	Same
	Library/Media Support	Same
	LEA Administrators	Same
	LEA Administrative Support	Same
	School Administrators	Same
	School Administrative Support	Same
	Student Support Services	Same
	All Other Support Services	Same