

# ASSESSING QUALITY OF CCD DATA USING A SCHOOL -BASED SAMPLE SURVEY

Sameena Salvucci, Sandeep Bhalla, Michael Chang, Synectics for Management Decisions, Inc. and  
John Sietsema, National Center for Education Statistics

Sameena Salvucci, Synectics for Management Decisions, Inc., 3030 Clarendon Blvd. #305, Arlington VA 22201

KEY WORDS: Administrative records, Sample surveys, Multiple reporting levels, Definitional differences, Record Matching

## I. Overview

This paper describes the methodology problems and data quality issues associated with the 1990-91 Common of Core of Data (CCD), a national database of universe data reflecting three levels of aggregation (state, local education agency, and school) collected from state education agency (SEA) administrative records. It evaluates the feasibility of using external school-based sample survey data, the 1990-91 Schools and Staffing Survey (SASS), to assess the accuracy of the CCD. It also describes the results of record matching procedures used to explain some of the existing differences between CCD and SASS.

## II. Context and Motivation

The National Center for Education Statistics (NCES) has been authorized by Congress, in part, to collect, analyze, and disseminate full and complete statistics on education in the United States. A primary way that NCES pursues this goal is through maintaining a comprehensive and timely national statistical database, the Common Core of Data (CCD). CCD is comprised of three separate nonfiscal surveys, the Public Elementary and Secondary School Universe (School Universe), the Local Education Agency Universe (LEA Universe), and the State Aggregate Nonfiscal Survey (State Aggregate). CCD provides general descriptive information, basic statistics, and fiscal data regarding all children in the United States enrolled in public schools, from prekindergarten through the twelfth grade, as well as staff, schools and local education agencies. However, participation in CCD is a voluntary activity of the states. NCES asks states to provide, from their administrative records, information they have secured from schools and LEAs. For the most part, the data requested by NCES are already collected by the states in the exercise of their responsibility for public education.

Given the variety among state definitions of the statistics being collected, there has been concern about how useful the national summaries of these data are. In

response, NCES has for many years conducted activities to develop standard definitions and procedures, help states observe these standards, and improve data quality of the CCD. Efforts continue to be devoted to improving CCD data accuracy. Since CCD has recently become the sampling frame for all NCES school-based surveys, new efforts are focusing on **measuring** the accuracy of the data. The work described in this paper, in particular, concentrates on measuring the accuracy of key statistical information in the CCD, such as the total number of students, teachers, schools and school districts.

While much can be learned from analysis of CCD data itself, another useful approach to measuring the accuracy of CCD is to compare it to data from other surveys. NCES's 1990-91 SASS, a national sample survey of public and private schools, is one such source of comparable data. SASS is comprised of four interrelated surveys. Three of these surveys are sent to public and private schools : (1) School Survey, (2) Administrator Survey, and (3) Teacher Survey and the fourth survey is sent to LEAs and is called the Teacher Demand and Shortage Survey (TDS).

## III. Consistency Within CCD

We began by examining the national estimates of student and FTE teacher counts from the three separate Nonfiscal surveys of CCD (School Universe, LEA Universe, and State Aggregate) for 1988-89, 1989-90, and 1990-91. The national level differences between the estimates from the three CCD survey components for a particular year showed improvement over time, and in 1990-91 no difference was larger than 1.2 percent. Even though the national level differences in 1990-91 were small, further examination identified some large differences between the three CCD components for some states. When student counts were summed up to the state level from the School Universe and compared to the student counts from the State Aggregate, two states had student count differences greater than 5 percent. When student counts from the School Universe were compared to the student counts from the LEA Universe, each aggregated to the state level, five states had student count differences greater than 10 percent. Even more striking, when FTE teacher counts were summed up to

the state level from the School Universe and compared to the FTE teacher counts from the State Aggregate, seven states had FTE teacher count differences greater than 20 percent.

Table 1 indicates the number of states where the student count differences exceeded one percent for the comparisons described above. The number of states with student count differences greater than one percent had discrepancies less than five percent consistently over the three years. The number of states with student count discrepancies greater than one percent decreased between 1988 and 1990, suggesting an overall improvement in the quality of CCD student count data over time.

Table 1 also lists the results of the FTE teacher count comparisons. Comparisons are not applicable between the school and LEA level for FTE teacher counts since the FTE teacher counts are not collected at the district level in CCD. Comparisons between the School Universe and the State Aggregate, however, do not show an improvement over the three years. In fact, the number of states with differences greater than 5 percent increased from 13 to 15. To add to the problem, some states have not been able to provide FTE teacher counts at the school level.

Across all the states, student count data exhibit more consistency across CCD survey components than FTE teacher counts, especially between the CCD School Universe and the State Aggregate Survey.

**Table 1:** Comparison of Estimates Aggregated to the State Level Between CCD Survey Components

Comparison	Number of States		
	1988-89	1989-90	1990-91
<b>Student Count (Difference &gt;= 1%)</b>			
School vs LEA	16	12	12
School vs State	13	8	9
<b>FTE Teacher Count (Difference &gt;= 5%)</b>			
School vs LEA	N/A	N/A	N/A
School vs State	13	12	15

Potential sources of discrepancies between these levels of reporting include:

- different interpretations by states of CCD definitions;
- variations in data collection and editing quality within states and in aggregating state reports to the national level; and
- external conditions that may limit the comparability of a reported item from state to state regardless of how well definitions and data processing standards are applied.

#### IV. CCD-SASS Consistency

The accuracy of the 1990-91 CCD estimates of interest were further examined by comparing this data to another survey, the 1990-91 SASS.

Both CCD and SASS provide estimates of student counts, FTE teacher counts, the number of schools and the number of school districts (LEAs). This section describes the multiple comparisons we used to assess the level of accuracy of these counts in CCD. Table 2 summarizes the eight different comparisons made.

**Table 2:** CCD-SASS Comparisons

1990-91 CCD	1990-91 SASS
<b>Student Counts</b>	
School Universe	Public School Survey
LEA Universe	TDS Survey
State Nonfiscal Survey	Public School
<b>FTE Teacher Counts</b>	
School Universe	Public School Survey
LEA Universe	TDS Survey
State Nonfiscal Survey	Public School Survey
<b>School Counts</b>	
School Universe	Public School Survey
<b>LEA Counts</b>	
LEA Universe	TDS Survey

For every comparison described in table 2 above, a two-step approach was used to make a decision on which states had the largest CCD-SASS differences for the estimates of interest. The first step identified the 95% confidence interval around the SASS estimate for

each state. Although the confidence intervals took sampling variance into account, some SASS estimates had very small standard errors or no standard error (e.g. states with only one LEA). The resulting confidence intervals for these SASS estimates had very small ranges which increased the likelihood that the corresponding CCD estimates would fall outside the interval. Therefore, we found that combining the confidence interval approach with examination of the actual percent difference between CCD and SASS was necessary. Examining the confidence intervals in conjunction with the percent difference provided a more realistic indication of large discrepancies.

The CCD state estimate was compared to the 95% confidence interval bounds around the SASS state estimate. For those states where the CCD estimate fell outside the 95% SASS confidence interval, the absolute value of the relative percent difference between CCD and SASS was calculated. When the CCD state estimate was both outside the 95% SASS state estimate confidence interval and the absolute value of the relative percent difference exceeded 10 then we identified the CCD state estimate as discrepant.

Table 3 lists the number of states where we identified the discrepancies between CCD and SASS as large according to the criteria described above. The intent was to define the extent of the problems requiring further investigation and to understand some of the issues surrounding cross-survey comparisons before any adjustments were made to the CCD data for differences in definitions between CCD and SASS.

**Table 3: Discrepant States Before Definition Adjustment**

Estimate	# of states where the estimate was identified as discrepant
<b>SASS School vs. CCD School</b>	
Schools	6
Students	2
FTE Teachers	10
<b>SASS TDS vs. CCD Agency</b>	
LEA's	11
Students	1
FTE Teachers	10
<b>SASS School vs. CCD Nonfiscal</b>	
Students	4
Teachers*	6

One possible source of the SASS-CCD discrepancies is that the definitions of what are nominally the same variables - LEA, school, student enrollment, and teacher count - substantially differ in their operational definitions between the SASS surveys and CCD. Recognizing that some of the discrepancies identified in table 3 were caused by these differences in definitions, we attempted to reconcile the data being compared.

The reconciliation of CCD data to SASS data was achieved through the creation of modified CCD files which more closely matched the SASS definitions of student enrollment count, FTE teacher count, number of schools and LEAs.

The process of producing revised estimates is described in two parts: steps common to all estimates, and steps devoted to producing revisions specific to the school, student, FTE teacher and the LEA counts.

SASS is only conducted in the 50 states and the District of Columbia (SASS Data File User's Manual 1990-91), whereas CCD consists of 50 states, the District of Columbia and the five U.S. outlying areas. As a result these outlying areas were removed for all the comparisons (Instructions for Completing the Nonfiscal Surveys of the Common Core of Data 1991). CCD defines a public school as an institution which provides educational services and has one or more grades prekindergarten through 12 or ungraded. On the other hand, SASS defines a school as an institution that provides educational services for at least one of grades 1 through 12 (or comparable ungraded levels). Since schools that offered only prekindergarten or kindergarten classes were not eligible for SASS, these schools were removed from CCD. A final adjustment was made by deleting all other schools in the 1990-91 CCD school universe which were not eligible for SASS.

In addition, prekindergarten enrollment was subtracted from the total student count for each school because SASS student enrollment at the school level includes only students from kindergarten through grade 12.

Adjustments were also made to compensate for the differences in the definition of teachers on CCD and SASS. On the SASS Public School Survey, a teacher is defined as any full-time or part-time teacher whose primary assignment was teaching in any of the grades kindergarten through grade 12. Itinerant teachers are included as part of the teacher count, as well as long-

term substitutes who were filling the role of a regular teacher on a long-term basis.<sup>1</sup> Short-term substitute teachers, student teachers, nonteaching specialists (e.g. guidance counselors and librarians), administrators, teacher's aides, and support staff are not included. These counts are head counts, NOT FTEs. In the CCD Public School Universe, however, the teacher count is stated in FTEs (full-time equivalents). This count includes only filled positions. The difference between a head count and a FTE teacher counts can be substantial. Also, CCD teacher counts include prekindergarten teachers, while SASS teacher counts do not. It is not possible to subtract the prekindergarten teachers from the total teacher counts as we did with the total student counts because CCD does not collect FTE teacher counts by grade level. Despite these problems we tried to match the two counts by converting the SASS head counts into FTE teacher counts. We created a derived FTE teacher count for SASS equal to the sum of the number of full-time teachers plus a weighted number of part-time teachers.

In SASS, an LEA is defined as a government agency that employ teachers. There are LEAs that employ teachers which do not operate schools. For example, some states have special education cooperatives that employ special education teachers who teach in one or more LEAs. In CCD, however, an LEA is defined as a government agency responsible for providing instructional services. The CCD definition does not mention teacher employment. In fact, the 1988-89 CCD frame included 1,352 LEAs which are not associated with schools, but hire teachers (Quality Profile for SASS, 4.2; The SASS Data File Users' Manual, p. 24). In order to include them in the SASS TDS population, a 1 in 10 systematic random sample of these districts was taken and included in the SASS sampling frame.

To replicate this design in our comparisons, all CCD districts linked to schools from the CCD school file were included in the comparisons. From this set we deleted all those LEAs which were only linked to those schools that had only prekindergarten or kindergarten enrollment. For those LEAs not linked to schools only a 1 in 10 sample was included (Quality Profile For SASS, p4.2).

---

<sup>1</sup>. An itinerant teacher is defined as a teacher who teaches at more than one school (for example, music teacher who teaches three days per week at one school and two days per week at another).

At the LEA level, the student count in CCD is reported as the sum of prekindergarten-12 plus ungraded students. To match the CCD enrollment figure with SASS School enrollment, the reported number of prekindergarten students for each school as provided in the CCD School Universe was aggregated for each LEA. The number of prekindergarten students was subtracted from the CCD LEA total student count.

For the state level comparisons, prekindergarten students are removed from the CCD State total student enrollment in order to derive a comparable estimate to the SASS survey.

The comparisons made for the FTE teacher counts followed the same adjustment as the school FTE count, but the data used was from the CCD State Aggregate Survey.

Table 4 lists the number of states where we identified the discrepancies between the reconciled CCD data and SASS data as large according to our two step criteria.

**Table 4 : Discrepant States After Adjustment**

Estimate	# of States where the estimate was discrepant
<b>SASS School vs. CCD School</b>	
Schools	3
Students	2
FTE Teachers	12
<b>SASS TDS vs. CCD Agency</b>	
LEA's	2
Students	0
FTE Teachers	3
<b>SASS School vs. CCD Nonfiscal</b>	
Students	2
Teachers	5

## V. Sources of CCD-SASS Inconsistency

After the identification of the set of states which have large CCD-SASS discrepancies for the estimates of interest, we focused our efforts on determining potential sources of the discrepancies within these states. Since SASS is a sample survey, we matched the set of all SASS schools and districts in the discrepant states with the corresponding CCD school and district in all three years 1988-89, 1989-90, 1990-91. We

limited our examination to only those districts and schools within the discrepant states for which the 1990-91 SASS-CCD discrepancy was greater than 10 percent for the student enrollment counts and greater than 20 percent for the FTE teacher counts. Next, we developed a flag that determined whether the size of the 1990-91 CCD-SASS discrepancy was larger than any discrepancy between the CCD counts across the three years examined. If the between year CCD differences were each smaller than the 1990-91 CCD-SASS difference, we felt that the 1990-91 CCD was probably more accurate than if the 1990-91 CCD-SASS difference was larger than any of the between year CCD differences for a particular school or LEA. In the latter case, we further compared the characteristics of these schools and districts as identified in both SASS and CCD. These comparisons indicated that most of the error was not due to random processing error, but rather showed systematic differences caused by different interpretations of definitions when reporting for SASS versus CCD. The following paragraphs describe some interesting findings.

The subsequent discussion is restricted to discrepancies in student counts and FTE teacher counts. Although CCD does not classify schools by level (elementary, secondary, combined), there should be consistency between the counts reported on CCD and the level reported on SASS. This is not always the case. For example, schools classified as secondary on SASS have no student counts in Grade 3, however there are 701 students reported for these schools on CCD. When SASS defined the school type as elementary, secondary or combined it was consistent with the definitions. That is, when the school was declared as elementary there was no student enrollment in SASS from grades 9 through 12, and when the school was secondary, there was no student enrollment in grade kindergarten through 6. On the other hand, CCD always reported student counts across the board even when the state was defined as elementary or secondary. Another interesting aspect of the analysis was that among the discrepant states across the eight comparisons the number of ungraded students reported always differed by a considerable amount.

We also conducted two rounds of interviews with state coordinators who submit the CCD data to NCES. We pinpointed states with large discrepancies between CCD and SASS in terms of student counts and FTE teacher counts. The first round of interviews resulted in a better understanding of which states were not able to collect certain types of information at certain levels and this was documented and helped reduce the

number of states we examined further. The second round of interviews elicited state coordinator input on more specific reasons for large discrepancies in school and district level discrepancies within the subset of states examined. Some very insightful details were recorded during the second set of interviews. CCD coordinators, in some cases, reported that State Education Departments were imputing these missing numbers, while others reported that the number was not correctly entered in the electronic version of the report. Some state coordinators reported that districts within their states were reporting teacher head counts instead of FTE teachers or were reporting FTE teacher counts for the LEA and state levels, while reporting teacher head counts at the school level. A major problem reported by states with large discrepancies was that they were double reporting the student counts. Specifically, one state coordinator explained that the schools in that state reported the vocational student counts, once as a part of the total student count and once as a part of the individual school count. As a result these numbers were counted twice in the total student population at the state. We also found during these interviews, that some states had students not enrolled in a school, but were enrolled in a district. As a result, districts in CCD were reporting a higher number of students compared to school level student counts when summed up to the district level. Errors were also reported when aggregating the LEA level enrollment in the CCD. Some states report the number of student counts in the LEA to a supervisory union. The schools also report their number to a supervisory union. The supervisory union in turn added the two counts and reported the sum (double count) as the LEA student count.

## VI. Next Steps

Many of the data quality assessment methods described in this paper could be used to further assess other important data elements in the 1990-91 CCD, such as the number of students by race/ethnicity. The quality of the 1993-94 CCD data could also be assessed using these methods, because survey data is now available for the 1993-94 SASS. Additional variables were collected in both the 1993-94 CCD, such as number of dropouts, that could be assessed using the Census Bureau's October School Enrollment Supplement to the Current Population Survey.

Electronic versions of this dataset are widely used by education researchers and policy makers. Any data quality problems at the school, district or state level could be included either in mainframe tape

documentation as technical notes and as context-sensitive help in the CCD CD-ROM.

------(1990-91), *Schools and Staffing Surveys: Data File User's Manual, Volume I*, U.S. Department of Education, Washington D.C.

## VII. Bibliography

Jabine T. (1992), *Quality Profile for SASS, Aspects of the Quality of Data in the Schools and Staffing Surveys*, U.S. Department of Education, Washington D.C.

Kaufman, S. And Huang, H.(1991), *Schools and Staffing Survey : Survey Design and Estimation*, NCES Technical Report, July 1993. U.S. Department of Education, Washington D.C.

National Center for Educational Statistics (1991-92), *Instructions for completing the Nonfiscal Surveys of the Common core of Data*, School Year 1991-1992, U.S. Department of Education, Washington D.C.

------(1989), *Data Base documentation, Common Core of Data Public School Universe, 1988-89*, U.S. Department of Education, Washington D.C.

------(1989), *Data Base documentation, Common Core of Data Public Education Universe, 1988-89*, U.S. Department of Education, Washington D.C.

------(1990), *Data Base documentation, Common Core of Data Public School Universe, 1989-90*, U.S. Department of Education, Washington D.C.

------(1990), *Data Base documentation, Common Core of Data Public Education Agency Universe, 1989-90*, U.S. Department of Education, Washington D.C.

------(1991), *Data Base documentation, Common Core of Data Public School Universe, 1990- 91*, U.S. Department of Education, Washington D.C.

------(1991), *Data Base documentation, Common Core of Data Public Education Universe, 1990-91*, U.S. Department of Education, Washington D.C.

------(1990-91), *SASS Data File User's Manual*, U.S. Department of Education, Washington D.C.