

SOME DATA ISSUES IN SCHOOL-BASED SURVEYS

Daniel Kasprzyk and Kerry Gruber, NCES, Sameena Salvucci, Mehrdad Saba, Fan Zhang,
and Steven Fink, Synectics for Management Decisions
[Sameena Salvucci, Synectics Mgmt Decisions, 3030 Clarendon Blvd 305, Arlington VA 22201]

KEY WORDS: Frame coverage; Response error adjustments; Sampling unit definitions

A. Introduction

The Schools and Staffing Survey (SASS) and the Teacher Follow-up Survey are periodic mail surveys conducted by the U.S. Bureau of the Census for the National Center for Education Statistics (NCES), U.S. Department of Education (Gruber, Rohr, Fondelier, 1993; Whitener et al., 1994).

At the National Center for Education Statistics (NCES), SASS is regarded as a major data set for providing information on teachers, principals, and schools. Its periodicity, three years between the first three rounds and now scheduled for four years between the third and fourth round of SASS, allows NCES the opportunity to investigate and study the consequences of decisions made in earlier rounds of the survey in preparation for the next data collection cycle.

During the last three years, the SASS program has initiated a number of projects aimed either at improving understanding of the SASS data or at clarifying a long-standing issue. This paper summarizes the results of three recent studies whose purposes originated with those goals. The concern of the first study was to evaluate how and whether changing the school sampling frame (and the definition of a school) affected SASS estimates. Some understanding of this issue can help in the interpretation of change estimates from Round 1 to Round 2.

The second study aimed to quantify the magnitude of an edit necessary to bring survey information as collected by the SASS in correspondence with frame information for an individual school, as obtained through the Common Core of Data (CCD), an annual NCES database with comparable statistical information for all public schools and school districts in the U.S. (McMillen, Kasprzyk, and Planchon, 1994). While there can be legitimate reasons for SASS and CCD to differ, large discrepancies from CCD are often indicative of problematic survey questions, survey procedures, or response error. Large differences between SASS and CCD had been observed for State estimates in ten states during data review prior to public release. These differences were reduced somewhat

through a post-processing edit (based on CCD data) of the individual school data for those ten states. This study extends the edit to the remaining 40 States and the District of Columbia and quantifies the changes in the estimates.

The third study identifies and compares estimates of the same or similar items across survey components. SASS has several built-in redundancies across its various components to allow researchers to use several components of SASS individually, thus eliminating processing steps. While such redundancies can be useful, they can also be confusing because estimates developed by researchers often differ, depending on the source of the data. The aim of the study was primarily to assist users and developers of SASS data to identify and understand differences in estimates of the same or similar items. The following sections describe the activities and results corresponding to the three studies.

B. Comparing SASS Estimates Using Different Sampling Unit Definitions

The public school sampling frame for the 1987-88 SASS was obtained from Quality Education Data, Inc. (QED). In this frame, a public school was defined as a physical unit or location. In the 1990-91 SASS, the public school sampling frame was based on the 1988-89 school year. The CCD-defined school is not a physical location, but an administrative unit. This difference in definition from the QED definition presented some concerns when the decision to change sampling frames was made. These concerns are well-founded, because some (CCD-defined) schools have two or more administrative units within one (QED-defined) physical location. This suggests that the estimates for the number of schools would be higher based on the CCD definition. The 1990-91 SASS sample design allows for the calculation of school, administrator, and teacher estimates using either the QED or the CCD definition of a school.

The purpose of this study was to measure the differences in estimates due to the difference in the CCD and QED definitions of a public school. Only 264 out of approximately 9,000 schools sampled in SASS were redefined. Knowing the extent of these differences and the characteristics of schools affected by

these definitional differences can guide the decision on how to make adjustments to the data for a trend analysis (Choy, Henke, Alt, Medrich, and Bobbitt, 1993) using the QED definition of school. Obtaining estimates based on the QED definition of school occurs by merging and identifying the multiple-CCD schools into the appropriate QED school, and summing the variables of interest across the CCD schools identified with the QED school. Weights for the QED schools are obtained by averaging all CCD schools' final weights within a QED-defined school.

Table 1 provides the QED- and CCD-defined estimates for the number of public schools and students for six states. These tables show the states most affected by the definitional change are North Dakota, South Dakota, Iowa, Nebraska, Minnesota, and Texas. This study showed only a small percentage of CCD-defined schools needed to be adjusted to meet the QED school definition. These schools, however, tended to be found in rural areas and states.

Table 2 provides the number of public schools and students by selected characteristics for rural/small towns and nationally under both definitions. The results showed more differences occur between the number of QED-defined schools and CCD-defined schools in small or rural towns versus urban fringe and large towns. The characteristics having the largest differences tend to occur as a result of the enrollment totals changing as two or more CCD schools are merged/defined as a QED school.

The most obvious ramification of this finding is that researchers analyzing rural trend data and some state trend data from the SASS need to be aware of the impact of these definitional differences on their analyses. For more details on this study see Holt and Scanlon (1994).

C. Effects of Post-Processing Edits on Survey Estimates

The initial review of the 1990-91 SASS data indicated the estimates of total teachers from the public school survey were at least 15 percent greater than the state Full-Time Equivalent (FTE) teacher counts reported on the 1990-91 CCD for nine states: Arkansas, Iowa, Missouri, Montana, Nebraska, North Dakota, Oklahoma, South Dakota, and Wisconsin; in addition, staff review of data from Arizona indicated data problems requiring further review (Gruber, Rohr, and Fondelier, 1993).

Two reasons were suggested for these overestimates. First, some schools did not appear to report data for their school but rather for their entire school district. At times this was due to vague or incorrect school names on the questionnaire label and at times the respondent misunderstood the instructions. The second factor contributing to the overestimates was that the survey respondents did not define schools in the same way that CCD did. For example, a school with grades K-8 at one address might be two CCD schools - an elementary school with grades K-6 and a middle school with grades 7 and 8; i.e., schools in SASS were reporting more grades than the same school had on the CCD (Gruber, Rohr, and Fondelier, 1993).

To make SASS state estimates of the number of teachers consistent with CCD, a post-processing edit was implemented to adjust the SASS data. The approach adopted was to edit SASS data to improve their consistency with CCD-reported data. The post-processing edit used the CCD school-level data for each school sampled in the 10 states to adjust the SASS data to CCD-appropriate grade ranges (Gruber, Rohr, and Fondelier, 1993) (table 3). The urgency to release the 1990-91 SASS data to the public precluded the NCES staff's ability to develop a comparable adjustment for the remaining 40 states and the District of Columbia. Thus, after the data were released a project was begun to develop a comparable adjustment and evaluate the impact of making adjustments to SASS estimates in the other 40 states. The principal concern with the released SASS data was the fact that the SASS data were processed differently in the two categories of states and that unknown biases existed in the data from the 40 states not included in the post-processing edit.

The study adjusted the 1990-91 SASS data to the appropriate CCD grade range following a set of decision rules intended to maintain the internal consistency of the reported data (Saba and Zhang, 1994), as was done with the ten states.

In comparing the CCD-adjusted and the original 1990-91 SASS estimates for FTE teachers (table 4) certain states stand out as being substantially affected by the CCD adjustment. The percent difference reflects the summed difference in SASS estimates and CCD-adjusted SASS estimates within each state.

D. Comparing Similar Estimates Across SASS Components

While the SASS survey is designed to be used across its school, district, administrator, and teacher components, researchers often conduct analyses using individual components. Reported results, therefore, would not usually uncover discrepancies from the same or similar survey items found in more than one component. Thus, the objectives of this study were to 1) identify and compare the same or similar survey items across the SASS and Teacher Follow-up Survey; and 2) compare national and state estimates for these items.

During the search for common variables across the surveys, attitudinal items were eliminated from the analysis. Results of this study are intended to assist researchers and users of the data to identify, help understand, and explain sources of variability on similar or the same survey items. They may also be of interest to persons responsible for various aspects of the design and operation of SASS.

After a review of the questionnaires, six variables were identified as being common on two or more surveys, including: school enrollment, teacher totals, teacher race/ethnicity, teacher certification, teacher training, and teacher attrition.

Public School K-12 Enrollment Comparisons.

This section compares the enrollment figures reported in SASS by school district administrators and principals. In the School District Survey, school district staff were asked to report student enrollment (in head counts) in six categories (ungraded, prekindergarten, kindergarten, grades 1-6, grades 7-12, and postsecondary), plus the total of these categories. Principals responding to the Public School Questionnaire were asked to report their student enrollment (in head counts) for each of the grade levels (16 categories) plus a total. Question wording and percentage distribution are located in figure 1.

Total K-12 enrollment. The first comparison examines enrollment estimates provided by LEAs and by the schools. Nationally, school estimates of total elementary and secondary enrollment are lower than district estimates by about one million students (or 2.5 percent). Examining total enrollment by state (not shown but available in Fink, 1994) reveals that school estimates are higher than district estimates in 19 states by an average of 2.9 percent and lower in 32 states by an average of 5.0 percent. There is a statistically significant difference between the district

and school enrollment estimates for 44 states. The District of Columbia shows the greatest difference with school totals almost 16 percent below district totals, followed by New Hampshire with district estimates greater than schools estimates by almost 11 percent.

Pre-Kindergarten enrollment. Nationally, pre-kindergarten enrollment estimates provided by schools are ten percent below district estimates (322,434 and 357,816, respectively). In 17 states, school estimates exceed district estimates by an average of 54 percent. In 32 states, school estimates are lower than district estimates by an average of 34 percent. In 11 states, the school estimates differ from the district estimates by more than 50 percent. Among the three states with the largest difference--Indiana, Montana, and Louisiana--school estimates are greater than twice the district estimates. All but seven states exceed the statistical significance level of .05. The detailed tables are available in Fink (1994).

Additional items were examined by Fink (1994). In general, estimates at the national level appear to differ by only a small percentage, though often being statistically significant. Comparing state estimates across SASS components often shows larger percentage differences. Individual categories, such as, ungraded, pre-kindergarten, and postsecondary also exhibit large differences across states.

Even though this study was initially aimed at assisting users of the SASS data, the most likely beneficiaries of the study are the data developers, who obviously must address serious conceptual and response issues for these items. Additional cognitive research, focus group research, pretesting, and user dialogue to determine the use of the various estimates in SASS is necessary.

Several reasons may account for the varying estimates from one survey to another. First, each component of SASS was completed by different respondents. The Teacher Demand and Shortage Survey was completed by school district personnel. Principals or headmasters/headmistresses completed the School Administrator Survey. The School Survey was completed by principals or individuals in the principal's office. Questions on The Teacher Survey were answered by currently employed school teachers. Finally, the Teacher Follow-up Survey questionnaires were sent a year later to a sample of participants in the SASS Teacher Survey. As a result, the quality of survey reports will differ.

Another reason why estimates on similar items may vary from one survey to another is the interview mode. SASS was designed to be primarily a mailout/mailback survey, but a substantial telephone follow-up was used for all sample units not returning the mail questionnaire (Jabine, 1994).

E. Endnote

The three studies summarized above provide an example of why data developers and data providers should try to maintain an inquisitive and questioning point of view. Each study aimed to provide a more thorough understanding of some aspect of the SASS data. Through these studies users can improve their understanding of the data they analyze, and data producers can take steps to improve the products they disseminate.

F. References

- Choy, S., Henke, R., Alt, M., Medrich, E., and Bobbitt, S. (1993), **Schools and Staffing in the United States: A Statistical Profile, 1990-91**. NCES 93-146. Washington, DC: U.S. Government Printing Office. Stock no. 065-000-00581-6.
- Fink, S. (1994), "1990-91 Schools and Staffing Surveys (SASS) Cross Questionnaire Analysis," NCES Working Paper, available from Special Surveys and Analysis Branch, NCES, Washington, D.C. 20208.
- Gruber, K., Rohr, C., and Fondelier, S. (1993), **1990-91 Schools and Staffing Survey: Data File User's Manual, Volume 1: Survey Documentation**, NCES 93-144. Office of Educational Research and Improvement, Washington, DC: National Center for Education Statistics.
- Holt, A. and Scanlon, B. (1994), "QED Estimates of the 1990-91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates," NCES Working Paper, available from Special Surveys and Analysis Branch, NCES, Washington, D.C. 20208.
- Jabine, T. (1994), **Quality Profile for SASS, Aspects of the Quality of Data in the Schools and Staffing Surveys (SASS)**, NCES 94-340. Office of Educational Research and Improvement, Washington, D.C.: National Center for Education Statistics.
- McMillen, M., Kasprzyk, D. and Planchon, P. (1993), "Sampling Frames at the United States National Center for Education Statistics," **Proceedings of the International Conference on Establishment Surveys**, Alexandria, VA: American Statistical Association.
- Saba, M., and Zhang, F., and Chang, M. (1994), "CCD Adjustment to the 1990-91 Schools and Staffing Survey (SASS)," NCES Working Paper, available from Special Surveys and Analysis Branch, NCES, Washington, D.C. 20208.
- Whitener, S., Rohr, C., Bynum, L., Kaufman, S., and King, K. (1994), **1991-92 Teacher Follow-up Survey: Data File User's Manual**. NCES 94-331 (Public Use Version) and NCES 94-478 (Restricted Use Version). U.S. Department of Education, Office of Educational Research and Improvement, Washington, DC: National Center for Education Statistics.

Table 1.--CCD and QED-defined estimates in SASS for number of public schools and students for selected states

	Schools		Students	
	CCD	QED	CCD	QED
U.S. Total	79,885	78,759	40,103,699	40,096,401
North Dakota	647	516	118,778	118,799
South Dakota	732	579	148,790	147,591
Iowa	1,530	1,445	479,023	478,912
Nebraska	1,455	1,325	260,030	260,211
Minnesota	1,434	1,346	719,581	719,460
Texas	5,651	5,606	3,323,523	3,323,498

Source: U.S. Department of Education, Schools and Staffing Survey: 1990-91 (School Questionnaire)

Table 2.--QED & CCD defined estimates for number of public schools and students, 1990-1991

	QED		CCD		Percent Difference	
	Schools	Students	Schools	Students	Schools	Students
U.S. Total	78,759	40,096,401	79,885	40,103,699	0.0	1.4
Rural/small town	39,263	15,694,730	40,352	15,695,586	2.8	0.0
School Level						
Elementary	25,715	9,395,915	26,508	9,495,515	3.3	0.0
Secondary	10,967	5,359,209	11,170	5,257,121	1.9	-1.9
Combined	2,581	939,606	2,674	942,951	3.6	0.4
Minority Enrollment						
Less than 20%	29,021	10,938,818	29,974	10,938,435	3.3	0.0
20% or more	10,242	4,755,912	10,378	4,757,151	1.3	0.0
School Size						
Less Than 150	6,938	594,261	7,843	664,432	13.0	11.8
150 to 499	21,179	6,700,298	21,477	6,746,207	1.4	0.7
500 to 749	7,304	4,418,856	7,252	4,383,991	-0.7	-0.8
750 or More	3,842	3,981,315	3,780	3,900,956	-1.6	-2.0

Source: U.S. Department of Education, NCES, Schools and Staffing Survey: 1990-91 (School Questionnaire)

Table 3.-- FTE Teachers for 1990-91 CCD and 1990-91 SASS After Adjustment (For Original 10 States)

State	CCD	SASS	SASS/CCD
U.S. Total	2,282,398	2,381,944	104.36%
Arizona	32,015	30,159	94.20%
Arkansas	25,787	27,091	105.06%
Iowa	31,795	33,402	105.05%
Missouri	51,115	52,632	102.97%
Montana	8,767	10,363	118.20%
Nebraska	18,771	18,107	96.46%
North Dakota	6,835	7,953	116.36%
Oklahoma	35,815	37,337	104.25%
South Dakota	8,389	9,863	117.57%
Wisconsin	50,724	55,207	108.84%

Source: Department of Education, NCES, 1990-91 CCD and 1990-91 SASS (School Questionnaire)

Note: All of the above states had a greater than 15 percent difference before adjustment.

Table 4.--FTE teachers for 1990-91 CCD, 1990-91 SASS Before and After CCD Adjustment

State	CCD	SASS		Percentage Effect of Adjustment
		Before Adjustment	After Adjustment	
U.S. Total	2,397,351	2,438,592	2,381,943	2.32%
Nevada	10,373	10,391	9,960	4.15%
Maine	15,513	16,069	15,289	4.85%
Louisiana	45,377	45,271	42,841	5.37%
Florida	108,088	105,167	99,479	5.41%
D.C.	5,950	5,543	5,956	7.45%
New Hampshire	10,637	10,852	9,924	8.55%
Minnesota	43,753	44,329	39,933	9.92%
Alaska	6,710	6,610	5,850	11.50%
Wyoming	6,784	7,349	6,151	16.30%

Source: U.S. Department of Education, NCES, 1990-91 CCD and 1990-91 SASS (School Questionnaire)

Figure 1.--Survey question wording, counts and percentage distributions

Question Wording	School District Survey Questionnaire: Question 1		Public School Survey Questionnaire: Question 17	
	What was the enrollment (in head counts) in this district on or about October 1 of THIS school year, and on or about October 1 of LAST school year?		How many students were enrolled in each grade on October 1 of this school year? (Report in head counts)	
Variables Used:	Counts	Distribution	Counts	Distribution
Ungraded	705,564	1.8%	321,721	0.8%
Kindergarten	3,237,854	7.9%	3,081,336	7.7%
Grades 1-6	19,419,747	47.5%	19,218,059	47.9%
Grades 7-12	17,482,583	42.8%	17,482,583	43.6%
Total	40,845,748	100.0%	40,103,699	100.0%

Source: NCES, Schools and Staffing Survey: 1990-1991 (School, District Questionnaire)