Technical Notes

FRSS 90: Internet Access in U.S. Public Schools, Fall 2005

Technical Notes — Internet Access in U.S. Public Schools: Fall 2005

Fast Response Survey System

The Fast Response Survey System (FRSS) was established in 1975 by the National Center for Education Statistics (NCES), U.S. Department of Education. FRSS is designed to collect issue-oriented data within a relatively short time frame. FRSS collects data from state education agencies, local education agencies, public and private elementary and secondary schools, public school teachers, and public libraries. To ensure minimal burden on respondents, the surveys are generally limited to three pages of questions, with a response burden of about 30 minutes per respondent. Sample sizes are relatively small (usually about 1,000 to 1,500 respondents per survey) so that data collection can be completed quickly. Data are weighted to produce national estimates of the sampled education sector. The sample size permits limited breakouts by classification variables. However, as the number of categories within the classification variables increases, the sample size within categories decreases, which results in larger sampling errors for the breakouts by classification variables.

Sample and Response Rates

The sample for the survey consisted of 1,205 regular elementary and secondary/combined schools in the 50 states and the District of Columbia. It was selected from the 2003–04 NCES Common Core of Data (CCD) Public School Universe file, which was the most current file available at the time of selection. The sampling frame included about 85,000 regular elementary and secondary/combined schools; about 63,000 of the schools were classified as elementary schools and about 21,000 were classified as secondary/combined schools. Special education, vocational education, and alternative/other schools were excluded from the sampling frame, along with schools with a highest grade below first grade and those outside the 50 states and the District of Columbia.

To select the sample, the frame of schools was stratified by instructional level (elementary, secondary/combined schools), enrollment size (less than 300 students, 300 to 999, 1,000 to 1,499, 1,500 or more), and percentage of students eligible for free or reduced-price lunch (less than 35 percent, 35 to 49 percent, 50 to 74 percent, 75 percent or more). Schools in the frame were then sorted by type of locale (city, urban fringe, town, and rural) and region (Northeast, Southeast, Central, and West) to induce additional implicit stratification.

All survey data were reported for fall 2005 when the survey was conducted. Questionnaires and cover letters for the study were mailed to the principal of each sampled school in early October 2005. Principals were told that the survey was designed to be completed by the technology coordinator or person most knowledgeable about Internet access at the school. Telephone follow-up for survey nonresponse and data clarification was initiated in November 2005 and completed in March 2006. During data collection, 34 schools were found to be ineligible for the survey because they had closed or merged with other schools. For the eligible institutions, the response rate was 86 percent (1,012 responding institutions divided by the 1,171 eligible schools in the sample). The weighted response rate was also 86 percent. Of the schools that completed the survey, 65 percent completed it by Web, 24 percent completed it by mail, 10 percent completed it by fax, and less than 1 percent completed it by telephone.

Although item nonresponse for key items was very low, missing data were imputed for the 14 items with a response rate of less than 100 percent. The missing items included both numerical data such as counts of instructional rooms and computers, as well as categorical data such as the provision of hand-held computers to students and teachers. The missing data were imputed using a "hot-deck" approach to obtain a "donor" school from which the imputed values were derived. Under the hot-deck approach, a donor school that matched selected characteristics of the school with missing data (the recipient school) was identified. The matching characteristics included level, enrollment size class, type of locale, and total number of computers in the school. Once a donor was found, it was used to obtain the imputed values for the school with missing data. For categorical items, the imputed value was simply the corresponding value from the donor school. For numerical items, an appropriate ratio (e.g., the proportion of instructional rooms with Internet access) was calculated for the donor school, and this ratio was applied to available data (e.g., reported number of instructional rooms) for the recipient school to obtain the corresponding imputed value. All missing items for a given school were imputed from the same donor. Imputation flags are included in the data.

Weighting Procedures and Sampling Errors

The response data were weighted to produce national estimates (see table 1). The weights were designed to adjust for the variable probabilities of selection and differential nonresponse. FRSS survey data are based on complex sample designs that require the use of weights to compensate for variable probabilities of selection, differential response rates, and possible deficiencies in the sampling frame. The reciprocal of the probability of selection, referred to as the "base weight," will produce unbiased (or consistent) estimates of population totals and ratios if there is no nonresponse in the survey. Since a stratified sample design was employed for the survey, the base weight for the *i*-th school in stratum *h* was computed as $w_{hi} = 1/f_h$ where f_h is the overall sampling rate used to select schools in stratum *h*.

Although the survey had a high response rate, adjustment of the base weights was necessary to compensate for the survey nonrespondents (i.e., whole questionnaire or unit nonresponse). To compensate for unit nonresponse, an adjustment factor was computed within selected weighting classes, and this factor was then used to inflate the base weights of the schools in the weighting class. The weighting classes used for this purpose were subsets of the sample defined by enrollment size class, poverty status, and type of locale. Within each subset, a nonresponse-adjustment factor was computed as the inverse of the base-weighted response rate. The factor was then applied to the base weights of the responding schools in the subset to obtain the final nonresponse-adjusted weight.

The survey findings were presented in a report titled *Internet Access in U.S. Public Schools and Classrooms: 1994–2005*. The reported findings are estimates based on the sample selected and, consequently, are subject to sampling variability. The standard error is a measure of the variability of an estimate due to sampling. It indicates the variability of a sample estimate that would be obtained from all possible samples of a given design and size. Standard errors are used as a measure of the precision expected from a particular sample. If all possible samples were surveyed under similar conditions, intervals of 1.96 standard errors below to 1.96 standard errors above a particular statistic would include the true population parameter being estimated in about 95 percent of the samples. This is a 95 percent confidence interval. For example, the estimated percentage of instructional rooms with Internet access in 2005 is 94 percent, and the estimated standard error is 1.3 percent. The 95 percent confidence interval for the statistics extends from 94 – (1.3 times 1.96) to 94 + (1.3 times 1.96), or from 91 to 97 percent. The coefficient of variation ("c.v.," also referred to as the "relative standard error") of an estimate (y) is defined as c.v. = (s.e. / y) x 100, where s.e. is the standard error of the estimate y.

Estimates of standard error were computed using a technique known as jackknife replication. As with any replication method, jackknife replication involves constructing a number of subsamples (replicates) from the full sample and computing the statistic of interest for each replicate. The mean square error of the replicate estimates around the full sample estimate provides an estimate of the variance of the statistic. To construct the replications, 50 stratified subsamples of the full sample were created and then dropped 1 at a time to define 50 jackknife replicates. A computer program (WesVar) was used to calculate the estimates of standard errors.

Table 1. Number and percent of responding public schools in the study sample, and estimated number and percent of public schools the sample represents, by school characteristics: 2005

School characteristic	Respondent sample (unweighted)		National estimate (weighted)	
	Number	Percent	Number	Percent
All public schools	1,012	100	82,482	100
Instructional level				
Elementary	489	48	61,923	75
Secondary	489	48	18,912	23
School size				
Less than 300	187	18	20,959	25
300 to 999	575	57	53,423	65
1,000 or more	250	25	8,100	10
Locale				
City	257	25	19,325	23
Urban fringe	339	34	27,637	34
Town	101	10	7,920	10
Rural	315	31	27,600	33
Percent minority enrollment				
Less than 6 percent	204	21	18,575	23
6 to 20 percent	245	25	21,043	26
21 to 49 percent	203	20	16,702	21
50 percent or more	340	34	24,192	30
Percent of students eligible for free or reduced-price lunch				
Less than 35 percent	395	39	32,281	39
35 to 49 percent	166	16	14,348	17
50 to 74 percent	217	21	18,289	22
75 percent or more	234	23	17,565	21

NOTE: Percent minority enrollment was not available for 20 schools. Thirty-four schools were combined schools and therefore are missing in the instructional level counts used here, but those cases were included in the totals and in analyses by other school characteristics. Detail may not sum to totals because of rounding or missing data.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Fast Response Survey System, "Internet Access in U.S. Public Schools, Fall 2005," FRSS 90, 2005.

Nonsampling Errors, Coding, and Editing

The survey estimates are also subject to nonsampling errors that can arise because of nonobservation (nonresponse or noncoverage) errors, errors of reporting, and errors made in data collection. These errors can sometimes bias the data. Nonsampling errors may include such problems as misrecording of responses; incorrect editing, coding, and data entry; differences related to the particular time the survey was conducted; or errors in data preparation. While general sampling theory can be used to determine how to estimate the sampling variability of a statistic, nonsampling errors are not easy to measure and, for measurement purposes, usually require that an experiment be conducted as part of the data collection procedures or that data external to the study be used.

To minimize the potential for nonsampling errors, the questionnaire was pretested with respondents at institutions like those that completed the survey. During the design of the survey and the survey pretest, an effort was made to check for consistency of interpretation of questions and to eliminate ambiguous items. The questionnaire and instructions were extensively reviewed by NCES.

Manual and machine editing of the questionnaire responses were conducted to check the data for accuracy and consistency. Cases with missing or inconsistent items were recontacted by telephone. A coding source file and editing manual were produced to identify cases requiring data retrieval or clarification and prepare cases for key entry. The source file served as a data dictionary and included the data file layout, a description of each data item, and a list of valid response codes, range formats, as well as codes for nonresponse, inapplicable responses, and defined skip patterns. The coding source file was used to develop the ACCESS database for data verification while the codebook served as the main tool for coding, editing, and processing questionnaires received by mail, fax, or telephone.

Logics, ranges, and validation checks were prepared prior to data collection and included online edit checks, manual logic checks, and automated checks using SAS. Online checks were incorporated into the online data entry system while manual edits were conducted to process cases received by mail, fax, or telephone. In both cases, where electronic and hardcopy survey data were submitted, steps were taken to ensure that the method of entering the data is the same, regardless of mode. For example, to enter survey data received by mail/fax or telephone, we accessed the survey website as "respondents" and "complete" the survey using the responses on the hardcopy survey. Subjecting all survey responses to the same set of built-in logics, ranges, and validation checks helps to ensure that data entry does not produce system differences in the survey data. In addition, all hardcopy data were subject to 100 percent verification and "double-keyed" in a simulated website interface for subsequent data checks.

Definitions of Selected Analysis Variables

Instructional level—Schools were classified according to their grade span in the 2003–04 Common Core of Data (CCD) Public Elementary/Secondary School Universe File. Data for combined schools are included in the totals and in analyses by other school characteristics, but are not shown separately. Thus, data are reported for the following categories:

Elementary school—Had grade 6 or lower and no grade higher than grade 8.

Secondary school—Had no grade lower than grade 7 and had grade 7 or higher.

School size—This variable indicates the total enrollment of students based on data from the 2003–04 CCD Public Elementary/Secondary School Universe File. For sampling purposes, schools were grouped into five enrollment size classes—less than 300 students, 300 to 499, 500 to 999, 1,000 to 1,499, 1,500 or more. Use of the more detailed size categories ensures greater diversity of schools in the sample with

respect to size, and permits a more nearly optimal allocation of the sample for estimating school-level characteristics that are correlated with enrollment. Because of the relatively small sample size and large standard errors associated with small cell sizes, the following three combined categories were used for analysis purposes:

Less than 300 students
300 to 999 students
1,000 or more students

Locale—This variable indicates the type of community in which the school is located, as defined in the 2003–04 CCD Public Elementary/Secondary School Universe File (which uses definitions based on U.S. Census Bureau classifications). The variable was based on the eight-category locale variable from CCD and collapsed into the following four categories for this report.

City—A central city of a Consolidated Metropolitan Statistical Area (CMSA) or Metropolitan Statistical Area (MSA).

Urban fringe—Any incorporated place, Census-designated place, or non-place territory within a CMSA or MSA of a large or mid-size city and defined as urban by the Census Bureau.

Town—An incorporated place or Census-designated place with a population greater than or equal to 2,500 and located outside a CMSA or MSA.

Rural—Any incorporated place, Census-designated place, or non-place territory designated as rural by the Census Bureau.

Percent minority enrollment—This variable indicates the percent of students enrolled in the school whose race or ethnicity is classified as one of the following: American Indian or Alaskan Native; Asian or Pacific Islander; Black, non-Hispanic; or Hispanic, based on data in the 2003–04 CCD Public Elementary/Secondary School Universe File. The categories are:

Less than 6 percent
6 to 20 percent
21 to 49 percent
50 percent or more

Percent of students eligible for free or reduced-price lunch—This variable was based on responses to question 25 on the survey questionnaire; if it was missing from the questionnaire (2.0 percent of all cases), it was obtained from the 2003–04 CCD Public Elementary/Secondary School Universe File. This item served as a measurement of the concentration of poverty at the school. The categories are:

Less than 35 percent
35 to 49 percent
50 to 74 percent
75 percent or more

Data Disclosure Warning

Under law, public use data collected and distributed by the National Center for Education Statistics (NCES) may be used only for statistical purposes.

Any effort to determine the identity of any reported case by public-use data users is prohibited by law. Violations are subject to Class E felony charges of a fine up to $250,000 and/or a prison term up to 5 years.

NCES does all it can to assure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are omitted or modified in the dataset to protect the true characteristics of individual cases. Any intentional identification or disclosure of a person or institution violates the assurances of confidentiality given to the providers of the information. Therefore, users shall:

- Use the data in this dataset for statistical purposes only.
- Make no use of the identity of any person or institution discovered inadvertently, and advise NCES of any such discovery.
- Not link this dataset with individually identifiable data from other NCES or non-NCES datasets.
- To proceed you must signify your agreement to comply with the above-stated statutorily based requirements.

Data perturbations were conducted on some background data to preclude identification of individuals and institutions.