Technical Notes
Distance Education Courses for Public Elementary and Secondary School Students, 2004–05

Data Disclosure Warning

Under law, public use data collected and distributed by the National Center for Education Statistics (NCES) within the Institute of Education Sciences 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.

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 timeframe. 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 FRSS survey on distance education courses in 2004–05 consisted of 2,312 public school districts in the 50 states and the District of Columbia. The nationally representative sample was selected from the 2003–04 NCES Common Core of Data (CCD) Local Education Agency Universe file, which was the most current file available at the time of selection. The sampling frame included 14,063 regular public school districts and 1,513 "other education agencies" with at least one charter school (referred to here as charter school districts). For the purposes of the study, "regular" school districts included any local school district that was not a component of a supervisory union (i.e., Education Agency type 1 on the CCD) or was a local school district component of a supervisory union sharing a superintendent and administrative services with other local school districts (i.e., Education Agency type 2 on the CCD). Also, charter school districts were "other education agencies" (i.e., districts with Education Agency type 7 on the CCD) that, when matched against the corresponding 2003–04 CCD Public School Universe file, had at least one charter school (i.e., had at least one school for which CHARTR03 = 1). Excluded from the sampling frame were districts in the outlying U.S. territories and regular districts with no enrollments or missing enrollments.[1] The sample of 2,312 districts included 2,211 regular school districts and 101 charter school districts. To allow for longitudinal analyses, the sample was designed to maximize overlap with the sample for the FRSS survey on distance education courses in 2002–03.[2] The general approach used was to assign conditional probabilities of selection to the eligible districts in the current sampling frame that depended on the desired selection probabilities for the current study and the selection probabilities for the previous study. The conditional probabilities were constructed in such a way that resulted in the desired (unbiased) selection probabilities for the current study while aiming to maximize overlap with the previous sample. In general, districts that were included in the previous sample and for which the desired selection probability under the current design was greater than or equal to the selection probability for the previous study were retained for the current sample. On the other hand, districts that were included in the previous sample and for which the desired selection probability under the current design was less than the selection probability for the previous study were given a chance of being deleted from the sample so as to maintain the desired overall sampling rates under the current design. The computational details of the sampling method are given in Brick, Morganstein, D., Wolters, C. (1987), "Additional uses for Keyfitz selection." *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, pp. 787-791, where it is shown that the procedure is unbiased.

Ninety-seven percent of the districts in the sample for the 2004–05 survey were also in the sample for the 2002–03 survey.[3] Although the study was designed primarily as a cross-sectional study, the use of the overlapping sample provided a longitudinal component that was used to simultaneously analyze responses from the two surveys, as reported in the Statistical Analysis Report, *Technology-Based Distance Education Courses for Public Elementary and Secondary School Students: 2002–03 and 2004–05* (NCES 2008–008). Such analyses required repeated measurements for the same districts that would not otherwise be possible with independent cross-sectional samples.

To select the sample, the school district sampling frame was stratified by district type (regular or charter), enrollment size (less than 1,000; 1,000 to 2,499; 2,500 to 9,999; 10,000 to 99,999; and 100,000 or more), and percentage of children in the district ages 5–17 in families living below the poverty level

[1] Charter school districts were included even if enrollment data were missing.

[2] The sample for the FRSS survey on distance education courses in 2002–03 used the same sample design and consisted of 2,305 public school districts in the 50 states and the District of Columbia. It was selected from the 2001–02 NCES Common Core of Data (CCD) Local Education Agency Universe file, which was the most current file available at the time of selection. The sampling frame included 14,229 regular public school districts and 989 "other education agencies" with at least one charter school.

[3] Of the 2,312 districts selected for the 2004–05 distance education survey, 2,242 districts had also been selected for the 2002–03 distance education survey.

(less than 10 percent, 10 to 19.99 percent, 20 to 29.99 percent, and 30 percent or more).[4] To improve the representativeness of the sample, an implicit stratification was induced by sorting the districts within each stratum by type of locale (city, urban fringe, town, rural)[5] and region (Northeast, Southeast, Central, West) prior to sampling.

Questionnaires and cover letters for the 2004–05 study were mailed to the superintendent of each sampled district in November 2005. The letter introduced the study and requested that the questionnaire be completed by the district's director of curriculum and instruction, the technology coordinator, the distance education coordinator, or another staff member who was most knowledgeable about the district's distance education courses. Respondents were offered the option of completing the survey via the web or by mail. Telephone follow-up for survey nonresponse was completed at the end of May 2006. Telephone follow-up for quality control and data clarification was completed in November 2006.

Of the 2,312 districts in the sample, 22 districts were found to be ineligible for the survey because the district had closed, merged with another district, or did not meet some other criteria for inclusion in the sample (e.g., they were administrative arms of a Board of Education). For the eligible districts, the response rate was 95 percent (2,176 responding districts divided by the 2,290 eligible school districts in the sample). The weighted response rate was 96 percent. Of the districts that completed the survey, 35 percent completed it by web, 38 percent completed it by mail, 10 percent completed it by fax, and 17 percent completed it by telephone.

Although item nonresponse for key items was very low, missing data were imputed for the 19 items with a response rate of less than 100 percent. The missing items included both numerical data, such as counts of enrollments in Advanced Placement courses offered through distance education, and categorical data, such as which technologies were used as primary modes of instructional delivery for distance education courses. Several questions contained multiple data items. These multiple items were imputed as a group to preserve their correlation. The missing data were imputed using a "hot-deck" approach to obtain a "donor" district from which the imputed values were derived. Under the hot-deck approach, a donor district that matched selected characteristics of the district with missing data (the recipient district) was identified. The matching characteristics included district type, region, metropolitan status, district enrollment size class, and poverty concentration. Once a donor was found, it was used to derive the imputed values for the district with missing data. For categorical items, the imputed value was simply the corresponding value from the donor district. For numerical items, the imputed value was calculated by taking the donor's response for that item (e.g., number of enrollments in advanced placement courses offered through distance education) and dividing that number by the total number of enrollments in distance education in the donor district. This ratio was then multiplied by the total number of enrollments in distance education in the recipient district to provide an imputed value. Imputation flags are included in the data.

[4]Poverty estimates for school districts were based on Title I data provided to the U.S. Department of Education by the U.S. Census Bureau and contained in U.S. Department of Commerce, U.S. Census Bureau, Current Population Survey (CPS) "Small Area Income and Poverty Estimates, Title I Eligibility Database, 2002." For detailed information on the methodology used to create these estimates, please refer to www.census.gov/hhes/www/saipe/index.html. The sampling categories were collapsed for analysis; see the section on Definitions of Selected Analysis Variables for more details.

[5]The 2003–04 CCD file contains two "urbanicity" variables: a three-category variable for metropolitan status (MSC03), and an eight-category variable for type of locale (LOCALE03). Type of locale was collapsed into four categories (city, urban fringe, town, and rural) and used for sampling. Metropolitan status was used in the Statistical Analysis Report and is included in the data file (with categories of urban, suburban, and rural); see the section on Definitions of Selected Analysis Variables for more details.

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 district in stratum h was computed as $w_{hi}=1/f_h$ where f_h is the overall sampling rate used to select districts 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 as the inverse of the base-weighted response rate within selected weighting classes. This factor was then used to inflate the base weights of the districts in the weighting class to obtain the final nonresponse-adjusted weight.

Table 1. Number and percent of public school districts in the study, and the estimated number and percent in the nation, for the total sample and for districts with students regularly enrolled in technology-based distance education courses, by district characteristics: 2004–05

	Total sample				Districts with students regularly enrolled in technology-based distance education courses			
	Respondent sample (unweighted)		National estimate (weighted)		Respondent sample (unweighted)		National estimate (weighted)	
District characteristic	Number	Percent	Number	Percent	Number	Percent	Number	Percent
All public school districts	2,176	100	15,190	100	802	100	5,670	100
District enrollment size								
Less than 2,500	1,013	47	11,120	74	365	46	4,150	74
2,500 to 9,999	740	34	3,090	21	238	30	1,070	19
10,000 or more	416	19	850	6	198	25	430	8
Metropolitan status								
Urban	296	14	1,530	10	109	14	380	7
Suburban	1,132	52	6,700	44	377	47	2,120	37
Rural	748	34	6,950	46	316	39	3,160	56
Region								
Northeast	466	21	2,910	19	113	14	630	11
Southeast	357	16	1,750	12	183	23	800	14
Central	700	32	5,650	37	283	35	2,550	45
West	653	30	4,880	32	223	28	1,690	30
Poverty concentration								
Less than 10 percent	842	41	5,210	38	293	37	1,840	34
10 to 19 percent	746	36	5,070	37	288	37	2,140	40
20 percent or more	485	23	3,330	24	204	26	1,440	27

NOTE: In the study sample, there were 7 cases for which district enrollment size was missing and 103 cases for which poverty concentration was missing. Detail may not sum to totals because of rounding or missing data for district characteristics. Poverty estimates for school districts were based on Title I data provided to the U.S. Department of Education by the U.S. Census Bureau.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Fast Response Survey System (FRSS), "Distance Education Courses for Public Elementary and Secondary School Students: 2004–05," FRSS 89, 2005.

The survey findings were presented in the Statistical Analysis Report, *Technology-Based Distance Education Courses for Public Elementary and Secondary School Students: 2002–03 and 2004–05* (NCES 2008–008). 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 public school districts with students regularly enrolled in distance education courses in 2004–05 is 37 percent and the standard error is 1.2 percent. The 95 percent confidence interval for the statistic extends from 37– (1.2 x 1.96) to 37 + (1.2 x 1.96), or from 34.6 to 39.4 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.

Because the data from the FRSS distance education courses survey were collected using a complex sampling design, the variances of the estimates from this survey (e.g., estimates of percentages) are typically different from what would be expected from data collected with a simple random sample. Not taking the complex sample design into account can lead to an underestimation of the standard errors associated with such estimates. Estimates of standard errors 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, 100 stratified subsamples of the full sample were created and then dropped one at a time to define 100 jackknife replicates. A computer program (WesVar) was used to calculate the estimates of standard errors using the JKN option.

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 error, the questionnaire was pretested with distance learning specialists, instructional technology specialists, or other people at the district who were deemed to be the most knowledgeable about the district's distance education courses. During the design of the survey and the survey pretest, an effort was made to check for consistency of interpretation of questions and definitions and to eliminate ambiguous items. The questionnaire and instructions were extensively reviewed by NCES and the data requester at the Office of Educational Technology.

Editing of the questionnaire responses was 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 specifications were used to produce the codebook. The codebook served as the main tool for coding, editing, and processing completed questionnaires. Coders used the codebook to identify cases

requiring data retrieval or clarification and prepare cases for entry into the web application. The source file served as a data dictionary and included the data file layout, a description of each data item, a list of valid response codes or range formats with codes for nonresponse and inapplicable, and defined skip patterns.

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 web application and manual edits were conducted to process cases received by mail, fax, or telephone. Steps were taken to ensure that the method of entering data from web and hardcopy questionnaires was the same, regardless of mode. For example, to enter survey data received by mail, fax, or telephone, the data processing staff accessed the survey website as "respondents" and "completed" 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 systematic differences in the survey data. In addition, all hardcopy data were subject to 100 percent verification using "doublekeying."

Definitions of Selected Analysis Variables

Many of the district characteristics, described below, may be related to each other. For example, district enrollment size and metropolitan status are related, with urban districts typically being larger than rural districts. Other relationships between these analysis variables may exist.

District Enrollment Size (SIZE)—This variable indicates the total number of students enrolled in the district based on data from the 2003–04 CCD. Data for this variable are missing for seven districts. The variable was collapsed into the following categories:

Not ascertained
Less than 2,500 students
2,500 to 9,999 students
10,000 or more students

Metropolitan Status (METRO)—This variable indicates the type of community in which the district is located, as defined in the 2003–04 CCD (which uses definitions based on U.S. Census Bureau classifications). Metropolitan status is the classification of an education agency's service area relative to a Metropolitan Statistical Area (MSA). An MSA is an area consisting of one or more contiguous counties (cities and towns in New England) that contain a core area with a large population nucleus, as well as adjacent communities having a high degree of economic and social integration with that core. An area is defined as an MSA if it is the only MSA in the immediate area and has a city of at least 50,000 population or it is an urbanized area of at least 50,000 with a total metropolitan population of at least 100,000 (75,000 in New England). The categories are described in more detail below.

Urban—Primarily serves a central city of an MSA

Suburban—Serves an MSA but not primarily its central city

Rural—Does not serve an MSA

Region (REGION)—This variable classifies districts into one of the four geographic regions used by the Bureau of Economic Analysis of the U.S. Department of Commerce and the National Assessment

of Educational Progress. Data were obtained from the 2003–04 CCD Local Education Agency Universe file. The geographic regions are as follows:

Northeast—Connecticut, Delaware, District of Columbia, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont

Southeast—Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia

Central—Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin

West—Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oklahoma, Oregon, Texas, Utah, Washington, and Wyoming

Poverty Concentration (POVERTY)—This variable indicates the percentage of children in the district ages 5–17 in families living below the poverty level, based on the Title I data provided to the U.S. Department of Education by the U.S. Census Bureau, "Small Area Income and Poverty Estimates." Data for this variable are missing for 103 districts. The variable was collapsed into the following categories:

Not ascertained
Less than 10 percent
10 to 19 percent
20 percent or more