Technical Notes – Dual Credit and Exam-Based Courses

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 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 FRSS survey on dual credit and exam-based courses consisted of 1,499 regular public secondary schools in the 50 states and the District of Columbia. It was selected from the 2001–02 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 17,059 regular secondary schools. For the purposes of the study, a secondary school was defined as a school with a grade 11 or 12. Special education, vocational education, and alternative/other schools were excluded from the sampling frame, along with schools with a highest grade below grade 11, those outside the 50 states and the District of Columbia, and schools with zero or missing enrollment.

To select the sample, the sampling frame was stratified by enrollment size (less than 300, 300 to 499, 500 to 999, 1,000 to 1,499, and 1,500 or more) and minority enrollment of the school (less than 6 percent, 6 to 20 percent, 21 to 49 percent, and 50 percent or more). Within each stratum, schools were sampled systematically and with equal probabilities at predetermined rates that varied from stratum to stratum. To improve the representativeness of the sample, an implicit stratification was induced by sorting the schools within each stratum by type of locale (city, urban fringe, town, rural) and region (Northeast, Southeast, Central, West) prior to sampling.

Questionnaires and cover letters for the study were mailed to the principal of each sampled school in mid-September 2003. The letter introduced the study and requested that the questionnaire be completed by the school's director of guidance counseling or other staff member who is most knowledgeable about the school's dual credit, Advanced Placement, and International Baccalaureate courses. Respondents were offered the option of completing the survey via the web or by mail. Telephone followup for survey nonresponse and data clarification was initiated in early October 2003 and completed in early January 2004.

Of the 1,499 schools in the sample, 32 were found to be ineligible for the survey because they did not have an 11th or 12th grade, were closed, or did not meet some other survey criteria (e.g., it was an alternative school). For the eligible schools, the response rate was 92 percent (1,353 responding schools divided by the 1,467 eligible schools in the sample). The weighted response rate was also 92 percent. Of the schools that completed the survey, 29 percent completed it by Web, 48 percent completed it by mail, 19 percent completed it by fax, and 5 percent completed it by telephone.

Although item nonresponse for key items was very low, missing data were imputed for the 39 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, as well as categorical data such as whether there were any requirements that students must meet in order to enroll in courses for dual credit. 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 enrollment size class and type of locale. 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, the imputed value was calculated by taking the donor's response for that item (e.g., enrollment in Advanced Placement courses) and dividing that number by the total number of students enrolled in the donor school. This ratio was then multiplied by the total number of students enrolled in the recipient school to provide an imputed value. All missing items for a given school were imputed from the same donor whenever possible. 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 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 schools in the weighting class to obtain the final nonresponse-adjusted weight.

The survey findings were presented in a report titled *Dual Credit and Exam-Based Courses in U.S. Public High Schools: 2002–03*. 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 high schools offering courses for dual credit is 71.3 percent, and the standard error is 1.4 percent. The 95 percent confidence interval for the statistic extends from 71.3 – (1.4 times 1.96) to 71.3 + (1.4 times 1.96), or from 68.6 to 74.0 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 dual credit and exam-based courses survey were collected using a complex sampling design, the variances of the estimates from this survey (e.g., estimates of proportions) 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, 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 using the JK1 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

Table 1. **Number and percent of responding public high schools in the study, and the estimated number and percent in the nation, for the total sample and for schools that offered courses for dual credit in 2002–03, by school characteristics: 2003**

School characteristic	Total sample				Offered courses for dual credit			
	Respondents (unweighted)		National estimates (weighted)		Respondents (unweighted)		National estimates (weighted)	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
All public high schools..................	1,353	100	16,480	100	1,016	100	11,750	100
Enrollment size								
Less than 500...................................	365	27	7,450	45	236	23	4,690	40
500 to 1,199....................................	470	35	4,960	30	355	35	3,730	32
1,200 or more	518	38	4,080	25	425	42	3,330	28
School locale								
City	269	21	2,740	17	193	19	1,790	15
Urban fringe	428	33	4,150	25	325	32	3,060	26
Town..	198	15	2,360	14	164	16	1,870	16
Rural	458	30	7,240	44	334	33	5,030	43
Percent minority enrollment								
Less than 6 percent..........................	398	30	5,640	35	316	32	4,290	37
6 to 20 percent.................................	335	25	3,770	23	269	27	2,950	25
21 to 49 percent...............................	290	22	3,170	20	223	22	2,300	20
50 percent or more	301	23	3,610	22	195	19	2,100	18

NOTE: Detail may not sum to totals because of rounding or missing data. For the FRSS study sample, there were 29 cases for which the percent minority enrollment in the school was missing. Those cases were included in the totals and in analyses by other school characteristics.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Fast Response Survey System (FRSS), "Dual Credit and Exam-Based Courses," FRSS 85, 2003.

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 directors of guidance counselors or other people at the school who were deemed to be the most knowledgeable about the school's dual credit, AP, and IB 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 Vocational and Adult Education.

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. The coding source file was used to develop the ACCESS database for data verification.

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 school characteristics, described below, may be related to each other. For example, school enrollment size and locale are related, with city schools typically being larger than rural schools. Other relationships between these analysis variables may exist.

Enrollment Size – This variable indicates the total number of students enrolled in the school based on data from the 2001–02 CCD. The variable was collapsed into the following three categories:

Less than 500 students (small)
500 to 1,199 students (medium)
1,200 or more students (large)

School locale – This variable indicates the type of community in which the school is located, as defined in the 2001–02 CCD (which uses definitions based on U.S. Census Bureau classifications). This variable was based on the eight-category locale variable from CCD and was recoded into a four-category analysis variable. Large and midsize cities were coded as city, the urban fringes of large and midsize cities were coded as urban fringe, large and small towns were coded as town, and rural areas outside and inside Metropolitan Statistical Areas (MSAs) were coded as rural. The categories are described in more detail below.

City – A large or midsize 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 midsize city, and defined as urban by the Census Bureau.

Town – Any 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 defined as rural by the Census Bureau.

Percent Minority Enrollment – This variable indicates the percentage of students enrolled in the school whose race or ethnicity is classified as one of the following: American Indian or Alaska Native, Asian or Pacific Islander, non-Hispanic Black, or Hispanic, based on data in the 2001–02 CCD School Universe file. Data on this variable were missing for 29 schools; schools with missing data were excluded from all analyses by percent minority enrollment. The percent minority enrollment variable was collapsed into the following four categories:

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