Technical Notes
Distance Education at Postsecondary Institutions, 2006-07

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.

Postsecondary Education Quick Information System

The Postsecondary Education Quick Information System (PEQIS) was established in 1991 by the National Center for Education Statistics (NCES), U.S. Department of Education (ED). PEQIS is designed to conduct brief surveys of postsecondary institutions or state higher education agencies on postsecondary education topics of national importance. Surveys are generally limited to 3 pages of questions, with a response burden of 30 to 45 minutes per respondent.

Most PEQIS institutional surveys use a previously recruited, nationally representative panel of institutions. The PEQIS panel was originally selected and recruited in 1991–92. In 1996, 2002, and 2006, the PEQIS panel was reselected to reflect changes in the postsecondary education universe that had occurred since the original panel was selected. A modified Keyfitz approach was used to maximize overlap between the panels for each reselection. This approach resulted in about 80 percent of the institutions overlapping for each reselection of the panel.

The sampling frame for the 2006 PEQIS panel was constructed from the 2005 Integrated Postsecondary Education Data System (IPEDS) Institutional Characteristics file. Institutions eligible for the 2006 PEQIS frame included 2-year and 4-year (including graduate-level) institutions that are both Title IV eligible and degree granting, and are located in the 50 states and the District of Columbia: a total of 4,265 institutions. The 2006 PEQIS sampling frame was stratified by instructional level (4-year, 2-year), control (public, private not-for-profit, private for-profit), highest level of offering (doctor's/first-professional, master's, bachelor's, less than bachelor's), and total enrollment. Within each of the strata, institutions were sorted by region (Northeast, Southeast, Central, and West) and by whether the institution had a relatively high minority enrollment. The sample of 1,627 institutions was allocated to the strata in proportion to the aggregate square root of total enrollment. Institutions within a stratum were sampled with equal probabilities of selection. Panel recruitment was conducted with the 339 institutions that were selected for the 2006 panel that were not part of the 2002 panel.

Each institution in the PEQIS panel was asked to identify a campus representative to serve as survey coordinator. The campus representative facilitates data collection by identifying the appropriate respondent for each survey and forwarding the questionnaire to that person. Data are weighted to produce national estimates, and the sample size allows for 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 distance education survey consisted of all of the institutions in the 2006 PEQIS panel. In addition, data were collected from one 4-year private for-profit institution that was added to the sample only for this survey because it is the largest provider of online distance education courses in the nation, bringing the total sample size for this survey to 1,628 institutions. In fall 2007, questionnaires were mailed to the PEQIS coordinators at the institutions. Coordinators were told that the survey was designed to be completed by the person at the institution most knowledgeable about the institution's distance education programs. Respondents had the option of completing the survey online. Telephone followup for survey nonresponse and data clarification was initiated 3 weeks after mailout and completed in March 2008.

Of the 1628 institutions in the sample, 18 were determined to be ineligible for the PEQIS distance education survey. These were mostly branch campuses of institutions for which distance education data were reported through a main campus. For the eligible institutions, the response rate was 90 percent (1,448 responding institutions divided by the 1,610 eligible institutions in the sample for this survey). The weighted response rate was 87 percent. Of the institutions that completed the survey, 72 percent completed it online, 20 percent completed it by mail, 5 percent completed it by fax, and 4 percent completed it by telephone.

Although item nonresponse for key items was very low, missing data were imputed for the items with a response rate of less than 100 percent. The missing items included both numerical data such as counts of students in online courses, as well as categorical data such as the extent to which institutions used various technologies for the instructional delivery of distance education courses. The missing data were imputed using a "hot-deck" approach to obtain a "donor" institution from which the imputed values were derived. Under the hot-deck approach, a donor institution that matched selected characteristics of the institution with missing data (the recipient institution) was identified. The matching characteristics included PEQIS stratum (defined by sector, highest level of offering, and enrollment size) and whether the institution offered any distance education courses. Once a donor was found, it was used to derive the

imputed values for the institution with missing data. For categorical items, the imputed value was simply the corresponding value from the donor institution. For numerical items, the imputed value was calculated by taking the donor's response for that item (e.g., enrollment in online courses) and dividing that number by the total number of students enrolled in the donor institution. This ratio was then multiplied by the total number of students enrolled in the recipient institution to provide an imputed value. All missing items for a given institution were imputed from the same donor whenever possible. Imputation flags are included in the data file.

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. PEQIS 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 institution in stratum *h* was computed as $w_{hi}=1/f_h$ where f_h is the overall sampling rate used to select institutions in stratum *h*.

Table 1. Number and percent of Title IV degree-granting postsecondary institutions in study, and estimated number and percent in the nation, for the total sample and for institutions that offered any distance education courses, by institutional type and institution size: 2006–07

	Total sample				Offered any distance education courses during the 2006–07 12-month academic year			
	Responding institutions (unweighted)		National estimate (weighted)		Responding institutions (unweighted)		National estimate (weighted)	
Institutional type and size	Number	Percent	Number	Percent	Number	Percent	Number	Percent
All institutions	1,448	100	4,200	100	1,179	100	2,750	100
Institutional type								
Public 2-year	509	35	1,000	25	502	43	1,020	37
Private not-for-profit 2-year	15	1	100	3	5	#	30	1
Private for-profit 2-year	65	4	500	12	12	1	90	3
Public 4-year	390	27	600	15	361	31	570	21
Private not-for-profit 4-year	419	29	1,500	36	262	22	790	29
Private for-profit 4-year	50	3	300	8	37	3	240	9
Size of institution								
Less than 3,000	511	35	2,700	65	294	25	1,400	51
3,000 to 9,999	487	34	900	23	449	38	870	32
10,000 or more	450	31	500	12	436	37	480	17

Rounds to zero.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Postsecondary Education Quick Information System (PEQIS), "Distance Education at Postsecondary Institutions," 2007.

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 institutions in the weighting class to obtain the nonresponse-adjusted weight.

The survey findings were presented in a *First Look* report titled *Distance Education at Degree-Granting Postsecondary Institutions: 2006-07* (NCES 2009-044). 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 Title IV degree-granting postsecondary institutions that offered any distance education courses is 65.9 percent and the standard error is 1.3 percent. The 95 percent confidence interval for the statistic extends from 65.9 – (1.3 x 1.96) to 65.9 + (1.3 x 1.96), or from 63.4 to 68.4 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 PEQIS distance education 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, 51 stratified subsamples of the full sample were created and then dropped 1 at a time to define 51 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 individuals at postsecondary institutions deemed to be the most knowledgeable about the distance education offerings at their institutions. 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 requestor 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

- **Institutional type (TYPE2):** public 2-year, private not-for-profit 2-year, private for-profit 2-year, public 4-year, private not-for-profit 4-year, private for-profit 4-year. This variable was created from a combination of level (2-year, 4-year) and control (public, private not-for-profit, private for-profit). Two-year institutions are defined as institutions at which the highest level of offering is at least 2 but less than 4 years (below the baccalaureate degree); 4-year institutions are those at which the highest level of offering is 4 or more years (baccalaureate or higher degree).[1]

- **Institution size (SIZE):** less than 3,000 students; 3,000 to 9,999 students; and 10,000 or more students.

[1]Definitions for level are from the data file documentation for the Integrated Postsecondary Education Data System (IPEDS) Institutional Characteristics file, U.S. Department of Education, National Center for Education Statistics.