

Technical Notes

Programs and Services for High School English Learners

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 (IES) 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 1,200 to 1,800 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 analysis variables. However, as the number of categories within any single analysis variable increases, the sample size within categories decreases, which results in larger sampling errors for the breakouts by analysis variables.

Sample Design and Selection

The sample for the FRSS survey *Programs and Services for High School English Learners* consisted of approximately 1,700 regular public school districts with high school grades in the 50 states and the District of Columbia. The nationally representative sample was selected from the 2012–13 NCES Common Core of Data (CCD) Local Education Agency Universe (LEA) file, which was the most current file available at the time of selection. The sampling frame included about 11,400 regular public schools districts that were coded with a high grade of instruction of 11 or 12 in the CCD LEA file. Regular school districts are generally administered by local education agencies and are responsible for providing instruction. They exclude supervisory unions that provide administrative services to multiple districts, regional education service agencies, state or federally

operated school districts, and charter school districts.¹ Excluded from the sampling frame were districts with a high grade of instruction below grade 11, districts with enrollment coded as zero, missing, “not applicable,” or “does not meet NCES quality standards,” districts that were not regular school districts, and districts outside the 50 states and the District of Columbia.

To select the sample, the district sampling frame was stratified by district enrollment size (less than 1,000; 1,000 to 2,499; 2,500 to 9,999; 10,000 to 24,999; 25,000 to 99,999; 100,000 or more) and percent of students in the district who are English learners (missing, less than 10 percent, 10 percent or more)² to create 17 primary strata. Within each stratum, districts were sorted by community type (city, suburban, town, and rural) and region (Northeast, Southeast, Central, and West) prior to selection to induce additional implicit stratification. The variables on the dataset for district enrollment size, community type, and region are defined in more detail in the “Definitions of Analysis Variables” section of these technical notes. Within each primary stratum, districts were selected systematically using sampling rates that depended on the size classification of the district.

Data Collection and Response Rates

Questionnaires and cover letters were mailed to the superintendent of each sampled district in September 2015. The letter stated the purpose of the study and requested that the questionnaire be completed by the person(s) in the district most knowledgeable about programs and services for English learners at the high school level. Respondents were asked to respond for the current 2015–16 school year. Respondents were offered options of completing the survey on paper or online. Telephone follow-up for survey nonresponse and data clarification was initiated in October 2015 and completed in February 2016.

Of the approximately 1,700 school districts in the sample, approximately 40 were found to be ineligible because the district was closed or did not meet some other criterion for inclusion in the sample (e.g., did not have high school grades). For the eligible districts, an unweighted response rate of 89 percent was obtained for this survey (about 1,480 responding districts divided by the approximately 1,670 eligible districts in the sample). The corresponding weighted response rate using the initial base weights was 89 percent. Among the respondents who completed the survey, 67 percent completed it via the Web, 28 percent completed it by paper (sent by mail, fax, or email), and 5 percent completed it by telephone.

Imputation for Item Nonresponse

Cases with missing data were recontacted by telephone to collect the missing information. However, for cases in which this data retrieval was unsuccessful, missing data were imputed. Although item nonresponse was very low (less than 1 percent for any item), missing data were imputed for the 32 items with a response rate of less than 100 percent. Table 1 shows the weighted percent of schools with imputed data for each questionnaire item. The missing items included both numerical data such as the number of high school English learners, as well as categorical data, such as whether the district had various programs or services designed specifically for English learners in high school. 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 enrollment size, community type, region, and percent of English

¹ For purposes of this study, “regular” school districts were those with TYPE equal to 1 or 2 on the CCD file (a local school district that is not a component of a supervisory union, or a local school district component of a supervisory union sharing a superintendent and administrative services).

² Percent of students in the district who are English learners (ELs) served as a measure of the concentration of ELs in the district. It was calculated by dividing the number of students in the district (at all grade levels) who were enrolled in programs for English language learners by the total number of students enrolled in the district at all grade levels; both counts were taken from the 2012–13 NCES Common Core of Data (CCD) Local Education Agency (LEA) Universe file.

learner students in the district. In addition, other relevant questionnaire items were used to form appropriate imputation groupings. Once a donor was found, the imputed value was simply the corresponding value from the donor district. Imputation flags are included in the data.

Table 1. Weighted percent of public school districts with imputed data, by questionnaire item: School year 2015–16

Questionnaire item		Percent imputed (weighted)
Question 2.	What is the current total number of high school English learners enrolled in your school district?	
Q2	Number of high school ELs enrolled in district	0.16
Question 3.	In your district, which of the following English learner instructional programs/approaches are currently provided for English learners in high school?	
Q3B	Two-way bilingual/dual language program in content classes	0.08
Q3F	Instructional support by a paraprofessional who speaks the student's native language	0.07
Q3G	Sheltered English/content instruction	0.04
Question 9.	In your district, do high school English learners work with online or computer-based programs in the following areas to address any of their needs as English learners?	
Q9C	ELs work with online or computer-based programs in content area instruction	0.13
Q9D	ELs work with online or computer-based programs for native language support in content area instruction	0.08
Q9E	ELs work with online or computer-based programs for organizational and study skills	0.13
Q9F	ELs work with online or computer-based programs in other areas	0.08
Question 10.	In your district, approximately how many high school English learners participate in the following programs and services?	
Q10B	How many ELs participate in remediation classes	0.20
Q10G	How many ELs participate in distance education course/program	0.08
Q10J	How many ELs participate in a mentoring program	0.08
Question 11.	Does your district have the following programs or services designed specifically for English learners in high school?	
Q11B	District has summer school designed specifically for ELs	0.04
Q11D	District has mentoring program designed specifically for ELs	0.04
Q11E	District has distance education course/program designed specifically for ELs	0.04
Q11F	District has other programs or services designed specifically for ELs	0.04
Question 13.	In your district, what is the approximate number of high school English learners with their native language used for content instruction (part 1), and what is the approximate number with their native language used for instructional support only (part 2)? In row a, provide information for the most common native language of high school ELs in the district. In row b, provide information for other non-English languages of high school ELs in the district.	
Q13APT1	Approximate number of HS ELs with the most common native language used for content instruction	0.13
Q13APT2	Approximate number of HS ELs with the most common native language used for instructional support only	0.21
Q13BPT1	Approximate number of HS ELs with another non-English language used for content instruction	0.13
Question 15.	Does your district provide English learners ages 18 to 21 seeking to newly enroll in your public school district with information about the following educational programs or services?	
Q15B	Provide ELs 18 to 21 seeking to newly enroll in the district with information about alternative school or program for at-risk students	0.08
Q15C	Provide ELs 18 to 21 seeking to newly enroll in the district with information about district-administered newcomer program	0.08

See notes at end of table.

Table 1. Weighted percent of public school districts with imputed data, by questionnaire item: School year 2015–16—Continued

Questionnaire item		Percent imputed (weighted)
Q15F	Provide ELs 18 to 21 seeking to newly enroll in the district with information about GED or adult education programs offered by the district	0.08
Q15G	Provide ELs 18 to 21 seeking to newly enroll in the district with information about GED or adult education programs offered by other entities	0.08
Q15H	Provide ELs 18 to 21 seeking to newly enroll in the district with information about free or low-cost English classes	0.08
Q15I	Provide ELs 18 to 21 seeking to newly enroll in the district with information about other educational programs or services	0.08
Question 16.	To what extent does your district consider the following factors when providing information about educational programs or services available to English learners ages 18 to 21 who are seeking to newly enroll in your school district?	
Q16A	Extent the district considers English proficiency level when providing info to ELs 18 to 21 who are seeking to newly enroll	0.08
Q16B	Extent the district considers literacy in their native language when providing info to ELs 18 to 21 who are seeking to newly enroll	0.08
Q16C	Extent the district considers limited or interrupted formal education when providing info to ELs 18 to 21 who are seeking to newly enroll	0.19
Q16D	Extent the district considers length of time needed to accrue sufficient credits to graduate when providing info to ELs 18 to 21 seeking to newly enroll	0.08
Q16E	Extent the district considers whether the student will be able to meet HS graduation requirements in content area classes when providing info to ELs 18 to 21 who are seeking to newly enroll	0.08
Q16F	Extent the district considers whether the student will be able to pass state tests required for graduation when providing info to ELs 18 to 21 who are seeking to newly enroll	0.15
Q16G	Extent the district considers age of student at time of enrollment when providing info to ELs 18 to 21 who are seeking to newly enroll	0.08
Q16H	Extent the district considers other factors when providing info to ELs 18 to 21 who are seeking to newly enroll	0.08

NOTE: Percentages are calculated as the weighted number of imputed cases divided by the weighted number of questionnaire respondents for whom the question applied. Only questionnaire items with missing data are listed in the table.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Fast Response Survey System (FRSS), “Programs and Services for High School English Learners,” FRSS 107, 2016.

Weighting Procedures and Sampling Errors

The responses were weighted to produce national estimates (see table 2). The weights were designed to reflect the probabilities of selection of the sampled districts and were adjusted for differential unit (questionnaire) 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). Nonresponse weighting adjustments were made within classes defined by district enrollment size and percent of the district enrollment in English language learner programs. Within the final weighting classes, the base weights (i.e., the reciprocal of districts’ probabilities of selection) of the responding districts were inflated by the inverse of the weighted response rate for the class.

Table 2. Number and percentage of responding public school districts in the study sample, and estimated number and percentage of public school districts the sample represents, by district characteristics: School year 2015–16

District characteristic	Respondent sample (unweighted)		National estimate (weighted) ¹	
	Number	Percent	Number	Percent
All public school districts	1,480	100	10,900	100
District enrollment size				
Less than 2,000	520	35	6,600	61
2,000 to 4,999	420	28	2,500	23
5,000 or more	540	36	1,800	16
Community type				
City	210	14	600	6
Suburban	470	32	2,300	22
Town	290	20	2,200	20
Rural	500	34	5,700	52
Region				
Northeast	290	20	2,200	20
Southeast	300	20	1,500	14
Central	450	31	4,200	38
West	430	29	3,000	28

¹ Weighted count of responding districts using the final nonresponse-adjusted weight (FWT). The weighted count is an estimate of the number of eligible districts in the study universe (see text for definition of the types of districts included in the study).

NOTE: Based on public school districts with high school grades. Detail may not sum to totals because of rounding.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Fast Response Survey System (FRSS), “Programs and Services for High School English Learners,” FRSS 107, 2016.

The survey findings are presented in a *First Look* report titled *Programs and Services for High School English Learners in Public School Districts: 2015–16* (NCES 2016-150). 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 percent of public school districts with high school English learners is 62 percent, and the standard error is 1.3 percent. The 95 percent confidence interval for the statistic extends from 62 – (1.3 x 1.96) to 62 + (1.3 x 1.96), or from 59.5 to 64.5 percent.

Because the survey data were collected using a complex sampling design, the variances of the estimates from the 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 under- or overestimation 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. A form of jackknife replication referred to as the JKN method was used to construct the replicates. Under the JKN method, the replicates were formed within groups of districts (called “variance strata”) within which districts were sampled at approximately the same rate. By creating the jackknife replicates within the variance strata, finite population correction factors (FPCs) can be introduced in the variance estimator to account for the fact that districts in some variance strata were sampled at relatively high rates. 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.

The sample of districts for this survey is relatively large compared to the population of eligible districts, so finite population correction (FPC) factors are required to estimate standard errors accurately; otherwise, the standard errors would tend to be overestimated. In addition to the FPC factors, a second set of factors referred to as JKN factors are also required to compute standard errors using the JKN option. The JKN factors pertain to the numbers of replicates that are formed for variance estimation. To facilitate loading of the factors into statistical software, the data for these factors are provided in the following two separate files: F107FACT_FPC.DAT is a text file containing the 100 FPC factors (one for each replicate), and F107FACT_JKN.DAT is a text file containing the corresponding 100 JKN factors. Alternatively, table 3 provides the same factor information contained in the text files.

Table 3. Values of finite population correction (FPC) and jackknife replication (JKN) factors to be used for calculating standard errors for FRSS 107: School year 2015–16

Replicate	FPC factor	JKN factor
1 to 10.....	0.45	0.900000
11 to 23.....	0.65	0.923077
24 to 59.....	0.81	0.972222
60 to 81.....	0.89	0.954545
82 to 100.....	1.00	0.947368

NOTE: FPC factors are based on the average sampling rate in the variance stratum.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Fast Response Survey System (FRSS), "Programs and Services for High School English Learners," FRSS 107, 2016.

Nonsampling Errors, Coding, and Editing

The survey estimates are 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 district-level personnel considered to be the most knowledgeable about programs and services for English learners at the high school level. 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 also extensively reviewed by NCES.

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 Analysis Variables

Many of the district characteristics, described below, may be related to each other. For example, district enrollment size and community type are related, with districts located in cities typically being larger than districts located in rural areas. Other relationships between these analysis variables may exist.

District enrollment size in three categories (N_DSIZCL)—This variable indicates the total number of students enrolled in the district based on data from the 2012–13 CCD Local Education Agency (LEA) Universe file. The variable used six categories for sampling, but was collapsed into the following three categories for the public use file:

Less than 2,000 students

2,000-4,999 students

5,000 or more students

Community type (URBAN)—This is a created variable collapsed from the 12-category urban-centric locale variable, as defined in the 2012–13 CCD Local Education Agency (LEA) Universe file. The urban-centric locale code is an indicator of a district’s location relative to a populous area. It is based upon the location of the school buildings in the district and in some cases may not reflect the entire attendance area or residences of enrolled students. This classification system has four major locale categories—city, suburban, town, and rural—each of which is subdivided into three subcategories. Community type was based on the 12-category urban-centric locale variable from CCD collapsed into the four categories below.

City—Territory inside an urbanized area and inside a principal city

Suburban—Territory outside a principal city and inside an urbanized area

Town—Territory inside an urban cluster

Rural—Territory outside an urbanized area and outside an urban cluster

Region (OEREG)—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. Data were obtained from the 2012–13 CCD Local Education Agency (LEA) 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