# CCD NONFISCAL UNIVERSE DATA REVIEWS AND EDITS

Contents

[CCD NONFISCAL UNIVERSE DATA REVIEWS AND EDITS 1](#_Toc65649952)

[Directory Review and Edits 1](#_Toc65649953)

[Directory Match edits 1](#_Toc65649954)

[Directory non-Match edits 1](#_Toc65649955)

[Data quality edits 1](#_Toc65649956)

[5-year edits 1](#_Toc65649957)

[Post-review Processing 2](#_Toc65649958)

[Suppressed Values 3](#_Toc65649959)

[Derived Variables 3](#_Toc65649960)

[Adjustments to State-Level Data 3](#_Toc65649961)

[Protecting Confidential Data 4](#_Toc65649962)

# CCD NONFISCAL UNIVERSE DATA REVIEWS AND EDITS

The public CCD data files undergo a significant data quality review prior to release. Additionally, in some cases NCES will make adjustments to SEA-level data to impute for missing data. The data quality edits, as well as the imputation process are described below.

## Directory Review and Edits

This group of edits are designed to ensure the core of the CCD data files, the Directory, is correct for all local education agencies (LEAs) and schools. These edits include Match and non-Match edits which are each described below.

### Directory Match edits

The first check of data submissions is to ensure that the reported universe of schools and LEAs is complete and accurate. To ensure the uniqueness of each reported entity and continuity with prior year reporting, the ESS runs a series of checks using the states’ own ID numbers for LEAs and schools, NCES ID numbers, and the reported status of the entities. These are referred to as “match” edits because they rely primarily on matching entities reported in the current year collection with those reported in the prior year. To pass these edits the reported status of the LEA or school must be logically consistent from prior year to current year (e.g., a school closed in the prior year should be reported as reopened, rather than open in the current year).

The companion file for directory data includes a cross tabulation of the counts of LEAs and schools by prior year and current year status with inconsistent status changes flagged. Closed schools and LEAs are reported on the CCD directory file for one year after they have closed but in subsequent years are not included in the CCD directory files.

### Directory non-Match edits

These edits include checks of LEA and school classifications, duplicate addresses, and phone numbers. Unlike the match edits described above, SEAs often resolve non-Match edits by providing an explanation or by confirming that the reported information is correct.

## Data quality edits

ED executes a series of data quality edits for all submitted data to flag missing or anomalous data for SEA review. These edits include checks to verify that details add to reported subtotals, data is reported for all operational entities, and membership counts reported by grade align with the grades offered for that entity.

States correct errors either by resubmitting the data or by providing a data note that explains why data that appears anomalous is actually correct.

## 5-year edits

NCES does not generally impute or adjust data at the LEA and school level. However, to avoid publishing clearly erroneous data, NCES tests for unlikely values and suppresses those that cannot be explained or resolved. Beginning with SY 2009–10, NCES implemented a review methodology that more accurately identifies data that are likely to be erroneous.

Prior to SY 2009–10, data elements were compared only to the reported data from the prior year for each school or LEA. The revised methodology uses available data from multiple prior years and uses both the reported values and derived ratios. Specifically, an average variation over the prior 4 years is computed based on the differences between each year and the other 3 years. Then, the average variation between the current year and the 4 preceding years is computed based on the differences between target year and each of the 4 prior years. Data elements with large differences are then flagged as potential errors. Additionally, a related ratio is calculated and tested. Examples of these computed ratios include pupil/teacher ratio and percent of students eligible for Free and Reduced-Price Lunches (FRPL). For example, in testing total student membership, the pupil/teacher ratio is also tested. A consistent value for the ratio indicates that the changed value is consistent with other changes in the data.

If a data element fails one of the multi-year edits it is suppressed due to data quality concerns if the state is unable to provide an explanation for the large change.

## Post-review Processing

NCES’ preference is to publish the data as reported by the SEAs. However, NCES makes some adjustments to the reported data for the following reasons:

* to set appropriate default values;
* to suppress enrollment or staff counts that are greatly inconsistent with prior year reports; this is done only if the SEA did not provide an adequate explanation for the anomaly;
* to add additional, derived variables;
* to impute for missing values in state-level data; and
* to suppress high or low counts that might compromise confidential data.

The field DMS\_FLAG in the companion files is used to indicate what, if any, adjustments were made to a submitted value. Most records will be flagged “Reported,” an indication that the data element is shown as originally submitted by the state. The following values of DMS\_FLAG are used in CCD data files:

* **Reported**. The data shown is as-reported by the state. The value for DMS\_FLAG in the vast majority of instances is “Reported.”
* **Suppressed**. The data have been suppressed due to concerns with the data quality.
* **Derived**. The record shown is derived from state-submitted data. Data users can often use these derived values as a shortcut in analyses as they provide common aggregations of data that states submit at a more granular level (e.g., teachers in an LEA or the total number of students by grade).
* **Missing**. The SEA may report a value as missing if they are unable to report the requested datum.
* **Not Reported**. The data for that item was not reported by the SEA. In creating the data extracts, records are added for category sets that the state did not report. If, for instance, a state does not submit records for secondary or elementary guidance counselors in an LEA, the value of DMS\_FLAG will be “Not Reported” for those records.
* **Not Applicable**. The SEA may report a value of “Not Applicable” for some data elements. For example, a state that does not have any grade 13 programs would report the counts of grade 13 as “Not Applicable”.
* **Post-submission edit**. The state submitted a value that has been changed by NCES analysts from the original submission. This is used rarely and done after notifying the state.
* **Manual adjustment**. The state-submitted values were adjusted (e.g., to conform with file specifications).
* **Imputation**. Data are imputed at the state level by NCES to inform national estimates. Individual imputations made by NCES are described in the “Post-submission data revisions” tab of the MS Excel file mentioned above.

### Suppressed Values

Values that failed the 5-year edit and were not resolved during data review are suppressed in post-review processing; the value is set to null in the long file and the DMS\_FLAG to ‘Suppressed’. Additionally, NCES may suppress data in other cases that is determined to be in error but these cases are rare.

### Derived Variables

Derived variables are a convenience to researchers and provide a consistent interpretation of the source data elements. The CCD nonfiscal universe data includes the following derived variables and subtotals:

SCH, LEA and SEA Directory Files:

* Grades Offered – Lowest (GSLO)
* Grades Offered – Highest (GSHI)
* LEA or school level (LEVEL)
* School and LEA Count (OPERATIONAL\_LEAS, OPERATIONAL\_SCHOOLS)

LEA and SEA Staff Files:

* Guidance Counselors, Teachers, School Staff, LEA Staff, and Other Staff

SCH, LEA, and SEA Membership Files:

* Derived - Education Unit Total minus Adult Education Count
* Derived - Subtotal by Race/Ethnicity and Sex minus Adult Education Count

## Adjustments to State-Level Data

In order to produce and publish state and nationally representative estimates for key data elements, NCES makes limited adjustments to state level CCD data.

Imputations are usually only done for missing data values. The particular imputation made is decided by NCES analysts on a case by case basis. Most often, a ratio is used to impute a given value. If, for example, a particular staff category was not reported in the current year, the prior year ratio of that staff count to total membership may be used to impute the missing staff count based on current year membership. The following paragraphs provide general descriptions of the types of adjustments made to CCD data.

#### Imputations from state’s own data

Where possible, NCES prefers to impute missing values using other data reported by the state. If data are missing for the current year, but were reported in the prior year, a current year value may be imputed using the prior year ratios of the state’s data. Depending on the item being imputed and the available data, these ratios could be of detail items to totals, staff counts to student membership, or a growth rate based on changes in student membership.

#### Imputations using averaged national ratio

If an imputation is not possible from the available data for a state, NCES may impute using averaged ratios of data from states that reported the necessary data. As with imputations that use a state’s own data, these averaged national ratios could be of detail items to totals, staff counts to student membership, or a growth rate based on changes in student membership.

#### Other Staff Categories Disaggregations

This adjustment is used in cases where the count for one staff category is not reported separately but is included within another category. It is pro-rated out at the SEA level. The approach used in these cases was as follows:

* Calculate the ratio of each missing item and the item containing the missing value(s) for each state reporting these items;
* Calculate the average of each of these ratios;
* Calculate the ratio of each average ratio to the sum of the average ratios for the unreported items as well as for the item that will be distributed; and
* Multiply the reported item by the ratio calculated in step 3 to determine the share for each item.

#### Ungraded teachers and students

SEAs provide additional information on data they report through the ED*Facts* EMAPS reporting system. Those data include flags to indicate whether the state uses “ungraded” (UG) as a designation for students or teachers as well as whether the state reports students enrolled in Grade 13. Starting with the SY 2015-16 CCD collection, adjustments were made to the reported numbers of Ungraded Students and FTE Ungraded Teachers, based on states’ EMAPS data. The following rules were applied.

* Retain all reported values (including zeros), even if the state indicated in EMAPS that they do not have ungraded students or ungraded staff.
* If UG data were expected (based on EMAPS) but not provided – the value is null and DMS\_FLAG=Not Reported.
* IF UG data were not expected (based on EMAPS) and not provided, that record is not included in the extract for that entity.

## Protecting Confidential Data

To protect against individuals being identified from CCD nonfiscal universe data, the counts of students with Individual Education Plans (IEP), English learners[[1]](#footnote-1), and students eligible for participation in the National School Lunch Program are subject to perturbation. This perturbation results in minimal changes to the data.

1. Starting with the SY 2019–20 reporting cycle, the English Learner data files will not be released with the CCD Universe Survey. Instead, these data will be released by the data steward through [ED Data Express](https://eddataexpress.ed.gov/). [↑](#footnote-ref-1)