

# State and National Estimation Data Notes

## **About the CRDC Data**

Since 1968, the Civil Rights Data Collection (CRDC) has collected data on key education and civil rights issues in our nation's public schools for use by the Department of Education's Office for Civil Rights (OCR), other Department offices, other federal agencies, and by policymakers and researchers outside of the Department. The CRDC has generally been collected biennially from school districts in each of the 50 states, and the District of Columbia. The CRDC collects information about school characteristics and about programs, services, and outcomes for students. Most student data are disaggregated by race/ethnicity, gender, limited English proficiency, and disability. The 2011-12 CRDC included all public schools and public school districts in the nation that serve students for at least 50% of the school day. The CRDC also includes long-term secure juvenile justice agencies, schools for the blind and deaf, and alternative schools.

## **About the State and National Estimations**

The national and state estimations presented in this report are based on data reported by the universe of U.S. public school districts. Data from school districts are self-reported. Despite being drawn from a universe of respondents, the submitted data may differ from their actual values due to the occurrence of non-sampling errors. Non-sampling errors are attributed to many sources, including definitional difficulties, the inability of respondents to provide accurate data, differences in the interpretation of questions, errors made in collection (e.g., in recording or coding the data), and errors made in estimating values for missing data. Quality control and edit procedures were used to reduce errors. However, users should carefully consider the caveats for analyzing the state and national estimations included in this document.

## **Race and Ethnicity**

School districts reported data using the seven race and ethnicity categories (American Indian or Alaska Native, Asian, Black or African American, Hispanic, Native Hawaiian or Other Pacific Islander, White, and Two or More Races). For more information on the Department's 2007 guidance regarding race and ethnicity categories, please visit <http://www2.ed.gov/policy/rschstat/guid/raceethnicity/index.html>.

## **Response Rates**

Traditionally the CRDC has achieved very high overall response rates. The response rate for this large national collection was 98.4% of school districts and 99.2% of schools, representing 99.6% of students in the nation.

An important consideration for response rates is that the reporting process for 2011-12 required all school and districts respond to each question on the CRDC prior to certification. Null or missing data prevented a school district from completing their CRDC submission to the Office for Civil Rights. Therefore, in cases where a school district may not have complete data, some schools or districts may have reported a zero value in place of a null value. It is not possible to determine all possible situations where this may have occurred. As such, it may be the case that the item response rates may be positively biased.

### **Suppression and Imputation**

Within the CRDC dataset, there are outliers that likely represented misreported values. These outliers had the potential to heavily influence state or national totals. To ensure the integrity of the state and national totals, OCR suppressed outliers identified by data quality rules. These rules flagged inconsistent and implausible values to be suppressed.

To mitigate the potential for suppressions that distort aggregate totals, suppressed data were replaced with imputed data where possible. For example, where the number of students disciplined exceeded the number in membership, the number was set to the number of students in membership.

The rate of imputation varied by data element. Data quality varied across states, so imputations occurred more frequently in some states than in others. The rate of imputation for bullying/harassment data was exceptionally high when compared to other data elements on the CRDC. For the large majority of CRDC data elements the rate of imputation ranged from 0-5% of reported values.

### **Privacy Protection**

To ensure the protection of privacy while meeting the purposes of the CRDC, state-level estimates were privacy protected by applying perturbation at the lowest level of disaggregation (for many data elements this is the cross-tabulation of race/ethnicity and sex). Perturbation, where applied, involved the addition or subtraction of 1 while preserving true zeros.

Given that the routine also incorporated data quality imputations, which introduced uncertainty into the data, a relatively low-frequency perturbation routine was used. Data values less than or equal to 10 had an *approximately* 5% likelihood of being perturbed. Data values greater than 10 had an *approximately* 1% likelihood of being perturbed.

### **Calculation of State and National Totals**

As mentioned, the privacy protection routine allows for the disclosure of true zeros for most CRDC data questions. For those questions, totals that are true zeros are published as zeros and counts of 1, 2, and 3 are published as “1-3” (bottom-coding). Bottom-coding was applied at the lowest level of disaggregated state-level data. For the CRDC questions that do require the protection of true zeros (Algebra I passing, AP exam passing, etc.), counts of 0, 1, 2, and 3 are published as “≤3”.

### **Data Considerations and Cautions**

Users should exercise caution when comparing the Part 1 number of students enrolled in Algebra 1 to the Part 2 number of students passing Algebra 1. The Part 1 data on the number of students enrolled in Algebra I was collected from the Fall of 2011, while the number of students passing Algebra I was collected at the end of the 2011-12 school year. Due to changes in enrollments over the school year, the count for enrollment and passing may not cover the exact same set of students.

Care must be taken in comparing estimations for various student populations because some of the difference may not be attributable to the education system. Some of the difference may be due to non-sampling error.

**Limitations of CRDC Data**

OCR strives to ensure CRDC data are an accurate and comprehensive depiction of student access to educational opportunities in school districts. The submission system includes a series of embedded edit checks to ensure significant data errors are corrected before the district submits its data. Additionally, each district is required to certify the accuracy of its submission. Only a district superintendent, or the superintendent's designee, may certify the CRDC submission. Ultimately, the quality of the CRDC data depends on accurate collection and reporting by the participating districts.

After reviewing the data, OCR is aware that inconsistencies may still remain in the data file. Users should be aware that outliers in the dataset may be a function of districts misreporting data. For example, outliers in the data on single-sex classes may be reporting the number of students enrolled in single-sex classes rather than the number of single-sex classes. In the analysis provided in this report, some schools and districts with potential reporting errors were excluded from the analysis. The percentage of schools included can be found in the notes section below each chart.

For more information about the CRDC, please visit:

<http://www2.ed.gov/about/offices/list/ocr/data.html>

**Contact information**

If you have any comments or questions concerning the use of CRDC data, please write to:

Office for Civil Rights

U.S. Department of Education

400 Maryland Avenue S.W.

Room 4E318

Washington, D.C. 20202-1172

Email: [ocrdata@ed.gov](mailto:ocrdata@ed.gov)