

## ***IDEA* Part B Data Collection**

### **Handling Missing Data When Reporting Race/Ethnicity**

(Revised January 2009)

This product is intended as a technical assistance document for data stakeholders who complete data submissions that include race/ethnicity information; it is consistent with *IDEA*, Section 618. The primary use of this product is for implementation of policy related to data collection and reporting. OSEP reviewed the document, and it is updated and distributed by the Data Accountability Center (DAC). Please direct your questions about this information to DAC at [IDEAdata@westat.com](mailto:IDEAdata@westat.com).

States should use the following guidelines only if they have NOT implemented the *Final Guidance on Maintaining, Collecting and Reporting Racial and Ethnic Data* to the U.S. Department of Education, published in the *Federal Register* October 2007 (Vol. 72, No. 202, available at <http://www.gpoaccess.gov/nara/index.html>). As described in the new guidance, the procedures for collecting, aggregating, and reporting race/ethnicity using seven (7) categories must be implemented by no later than the report of the SY 2010-11 *IDEA* data. Although not required, states may, as early as for reports referencing reporting period 2008-09, collect and aggregate their *IDEA* data as specified in the new guidance. When implementing the new guidance, states must do so at the beginning of the relevant reporting period and must apply the guidance consistently to all *IDEA* data reported within that period.

States are required to report race/ethnicity data for all students reported for the child count (Table 1), educational environments (Table 3), exiting (Table 4), and discipline (Table 5) data collections. For those states that are approved for *EDFacts*-only submission, the file specifications that align with the above tables are N002 and N089 (child count and educational environments); N009 (exiting); and N005, N006, N007, N088, N143, and N144 (discipline). There may be occasions, however, when the racial/ethnic background of some students is unknown. If a state has NOT implemented the new seven (7) race/ethnicity categories, it should use the following guidance for missing race/ethnicity data.

One method for addressing missing race/ethnicity data is to apply the racial/ethnic distribution of those students whose race/ethnicity is known to the students whose race/ethnicity is unknown. This method is described in more detail below.

Please note that OSEP is not asking states to assign a race/ethnicity to individual students. Rather, states are asked to estimate missing racial/ethnic data for aggregate data reporting only. When using this estimation method, states should provide a comment to describe the number of students in each reporting category for whom race/ethnicity was estimated.

The second method for addressing missing race/ethnicity is to omit children/students for whom race/ethnicity is unknown or who have multiple race/ethnicity status from the race/ethnicity cross-tabulations. Note that while a State may opt to omit children/students from the race/ethnicity cross-tabulations, they must be included in the row and column totals. This practice will produce red cell errors and requires a Data Note to explain the omission (including a breakdown of the number of omitted children/students for whom race/ethnicity is unknown vs. those who are multiracial).

#### **Levels of Data Reporting**

States should estimate race/ethnicity where the missing data occur in terms of: (1) reporting level, (2) age group, and (3) reporting category.

- (1) Reporting level. States should estimate missing racial/ethnic data at the school level whenever possible. When estimating missing school-level race/ethnicity data, use the school-level racial/ethnic distribution of the reporting category. If school-level data are unavailable, then estimate race/ethnicity at the district level using the district-level racial/ethnic distribution.
- (2) Age group. When estimating race/ethnicity, use the racial/ethnic distribution of the appropriate age group. Address missing data for 3- through 5-year-olds using the racial/ethnic distribution for children ages 3 through 5. Address missing data for 6- through 21-year-olds using the racial/ethnic distribution for students ages 6 through 21.
- (3) Reporting category. Estimate race/ethnicity using the racial/ethnic distribution of the reporting category with the missing data. For example, if missing race/ethnicity data occur in one of the disability categories, then use the racial/ethnic distribution of that specific disability category (e.g., mental retardation). If missing data occur in one of the educational environments, then use the racial/ethnic distribution of that specific environment (e.g., inside the regular class 80% or more of day). For exiting, use the racial/ethnic distribution of the specific basis for exiting, and for discipline use the racial/ethnic distribution of the specific discipline category.

### **Multiracial Children and Youth**

For multiracial children and youth, these estimation procedures are not ideal. The race/ethnicity of these students is known, so, when possible, report them in one of the five racial/ethnic categories based on their racial/ethnic background. Only estimate race/ethnicity for multiracial students if individual race/ethnicity data are not available (e.g., an aggregate collection that uses a single multiracial category).

### **Degree of Missing Data**

Before estimating race/ethnicity, states should investigate the degree of missing racial/ethnic data. This estimation method assumes that the racial/ethnic distribution of the students whose race/ethnicity is unknown is the same as the racial/ethnic distribution of the students whose race/ethnicity is known. This assumption may or may not be true. The larger the percentage of students with missing race/ethnicity data, the less likely it is this assumption is appropriate. As a result, inaccurate racial/ethnic estimates may occur. Whenever there is a large percentage of missing data, states should investigate why the data are missing and take appropriate action to correct the reporting problem.

### **Example #1: Child Count Data**

States report child count data by age and disability, as well as by race/ethnicity and disability. For each disability category, the number of students reported for the five racial/ethnic groups should sum to the total number of students reported by age for that disability category. For ages 3 through 5, this means that the disability totals in Section B-2 should equal the totals in Section B-1. For ages 6 through 21, this means that the disability totals for Section D should equal the 6-21 total column of Section C.

The following table shows fictitious race/ethnicity data for the students reported in three of the disability categories. In each example, the sum across racial/ethnic groups does not equal the total reported (by age) for the disability category. As shown in the last column, the racial/ethnic data are missing for several students.

Disability	American Indian/Alaska Native	Asian/Pacific Islander	Black (not Hispanic)	Hispanic	White (not Hispanic)	Racial/Ethnic Sum	Reported Total for Age Group	Unknown Race/Ethnicity
MR	0	20	50	50	80	200	225	25
ED	5	5	25	15	50	100	120	20
SLD	0	0	200	100	100	400	430	30

To estimate race/ethnicity for these students, calculate the racial/ethnic distribution of the disability category based on the students in the category whose race/ethnicity is known.

For example, for MR, there are 200 students whose race/ethnicity is known. None of them are American Indian/Alaska Native, 20 are Asian/Pacific Islander, 50 are Black, 50 are Hispanic, and 80 are White. Translating the counts into percentages, 0% of the 200 students are American Indian/Alaska Native (0/200), 10% are Asian/Pacific Islander (20/200), 25% are Black (50/200), 25% are Hispanic (50/200), and 40% are White (80/200). The known racial/ethnic distributions for the three disability categories are presented below.

Disability	American Indian/Alaska Native	Asian/Pacific Islander	Black (not Hispanic)	Hispanic	White (not Hispanic)
MR	0%	10%	25%	25%	40%
ED	5%	5%	25%	15%	50%
SLD	0%	0%	50%	25%	25%

Now, apply the racial/ethnic distribution to the students whose race/ethnicity is unknown. To do this, multiply the number of students whose race/ethnicity is unknown by each of the racial/ethnic group percentages to determine how many of those students to assign to each race/ethnicity group. Below are the examples.

<b>MR</b>		
American Indian/Alaska Native	$25 * 0.00$	0 students
Asian/Pacific Islander	$25 * 0.10$	3 students
Black (not Hispanic)	$25 * 0.25$	6 students
Hispanic	$25 * 0.25$	6 students
White (Not Hispanic)	$25 * 0.40$	10 students
<i>Total = 25 students</i>		

ED		
American Indian/Alaska Native	20 * 0.05	1 student
Asian/Pacific Islander	20 * 0.05	1 student
Black (not Hispanic)	20 * 0.25	5 students
Hispanic	20 * 0.15	3 students
White (Not Hispanic)	20 * 0.50	10 students
<i>Total = 20 students</i>		

SLD		
American Indian/Alaska Native	30 * 0.00	0 students
Asian/Pacific Islander	30 * 0.00	0 students
Black (not Hispanic)	30 * 0.50	15 students
Hispanic	30 * 0.25	8 students
White (Not Hispanic)	30 * 0.25	7 students
<i>Total = 30 students</i>		

Sometimes, due to rounding, the resulting number of students is one too few or one too many. This minor discrepancy can be easily adjusted. For example, if you have one too few students, identify the preredounded value with the largest decimal below 0.5 and round that count up. If you have one too many students, identify the preredounded value with the smallest decimal 0.5 or above and do not round that count up.

At the end of this step, you know how many of the unidentified students to assign to each racial/ethnic group. Add these students to the child count as reported by race/ethnicity data (see below).

Disability	American Indian/Alaska Native	Asian/Pacific Islander	Black (not Hispanic)	Hispanic	White (not Hispanic)	Racial/Ethnic Sum	Reported Total for Age Group	Unknown Race/Ethnicity
MR	0+0 = 0	20+3 = 23	50+6 = 56	50+6 = 56	80+10 = 90	225	225	0
ED	5+1 = 6	5+1 = 6	25+5 = 30	15+3 = 18	50+10 = 60	120	120	0
SLD	0+0 = 0	0+0 = 0	200+15 = 215	100+8 = 108	100+7 = 107	430	430	0

### **Example #2: Non-Child Count Data**

This estimation method can also be applied to non-child count data, including educational environments (Table 3), exiting (Table 4), and discipline (Table 5) data. For each data collection, the number of students reported for the five racial/ethnic groups should sum to the total number of students reported by age for that table. Any discrepancies indicate missing racial/ethnic data.

As discussed earlier, you should estimate missing race/ethnicity data using the known racial/ethnic distribution specific to the reporting category, reporting level, and age group where the missing data

occur. Use the distribution of the educational environment for educational environment data, the distribution of the exiting basis for exiting data, and the distribution of the discipline category for discipline data. Apply the estimating procedures to school-level data whenever possible.

The following table shows fictitious race/ethnicity data for the children reported in three of the 6-21 educational environments. In each example, the sum across racial/ethnic groups does not equal the total reported (by age) for the environment. As shown in the last column, the racial/ethnic data are missing for several children.

Educational Environment	American Indian/Alaska Native	Asian/Pacific Islander	Black (not Hispanic)	Hispanic	White (not Hispanic)	Racial/Ethnic Sum	Reported Total for Age Group	Unknown Race/Ethnicity
In the Regular Class 80% or More of the Day	25	50	200	125	100	500	550	50
Separate School	9	6	60	135	90	300	325	25
Residential Facility	28	12	100	40	220	400	440	40

To estimate race/ethnicity for these children, calculate the racial/ethnic distribution of the educational environment based on the children in the environment whose race/ethnicity is known.

For example, for the in the regular class 80% or more of the day category, there are 500 students whose race/ethnicity is known. Twenty-five of them are American Indian/Alaska Native, 50 are Asian/Pacific Islander, 200 are Black, 125 are Hispanic, and 100 are White. Translating the counts into percentages, 5% of the 500 children are American Indian/Alaska Native (25/500), 10% are Asian/Pacific Islander (50/500), 40% are Black (200/500), 25% are Hispanic (125/500), and 20% are White (100/500). The known racial/ethnic distributions for the three educational environments are presented below.

Educational Environment	American Indian/Alaska Native	Asian/Pacific Islander	Black (not Hispanic)	Hispanic	White (not Hispanic)
In the Regular Class 80% or More of the Day	5%	10%	40%	25%	20%
Separate School	3%	2%	20%	45%	30%
Residential Facility	7%	3%	25%	10%	55%

Now, apply the racial/ethnic distribution to the children whose race/ethnicity is unknown. To do this, multiply the number of children whose race/ethnicity is unknown by each of the racial/ethnic group percentages to determine how many of those children to assign to each racial/ethnic group.

In the regular class 80% or more of the day		
American Indian/Alaska Native	50 * 0.05	2 children
Asian/Pacific Islander	50 * 0.10	5 children
Black (not Hispanic)	50 * 0.40	20 children
Hispanic	50 * 0.25	13 children
White (Not Hispanic)	50 * 0.20	10 children
<i>Total = 50 children</i>		

Separate School		
American Indian/Alaska Native	25 * 0.03	1 child
Asian/Pacific Islander	25 * 0.02	0 children
Black (not Hispanic)	25 * 0.20	5 children
Hispanic	25 * 0.45	11 children
White (Not Hispanic)	25 * 0.30	8 children
<i>Total = 25 children</i>		

Residential Facility		
American Indian/Alaska Native	40 * 0.07	3 children
Asian/Pacific Islander	40 * 0.03	1 child
Black (not Hispanic)	40 * 0.25	10 children
Hispanic	40 * 0.10	4 children
White (Not Hispanic)	40 * 0.55	22 children
<i>Total = 40 children</i>		

At the end of this step, you know how many of the unidentified children to assign to each racial/ethnic group. Add these children to the educational environment as reported by race/ethnicity data (see below).

Educational Environment	American Indian/Alaska Native	Asian/Pacific Islander	Black (not Hispanic)	Hispanic	White (not Hispanic)	Racial/Ethnic Sum	Reported Total for Age Group	Unknown Race/Ethnicity
In the regular class 80% or more of the day	25+2=27	50+5=55	200+20=220	125+13=138	100+10=110	550	550	0
Separate School	9+1=10	6+0=6	60+5=65	135+11=146	90+8=98	325	325	0
Residential Facility	28+3=31	12+1=13	100+10=110	40+4=44	220+22=242	440	440	0

The same estimation method can also be applied to missing race/ethnicity data in the exiting and discipline data collections.