

## Appendix D: Data Quality Calculations

This appendix provides additional detail on how to conduct calculations to check the quality of the data. This information corresponds to [Section III Data Quality](#).

### Subtotal to Total

Below are the subgroups expected to equal the unduplicated count of students or another aggregate group. These examples refer to file specs 119 (SA participation) and 127 (LEA participation).

Subgroup	Subgroup Categories	Equation	Total
<b>Gender</b>	Male + Female	Equal to	Unduplicated count of students
<b>Ethnicity</b>	Non-Hispanic + Hispanic	Equal to	Unduplicated count of students
<b>Age</b>	Age 3 + Age 4 + Age 5 + Age 6 + Age 7 + Age 8 + Age 9 + Age 10 + Age 11 + Age 12 + Age 13 + Age 14 + Age 15 + Age 16 + Age 17 + Age 18 + Age 19 + Age 20 + Age 21	Equal to	Unduplicated count of students
<b>Race</b>	Native American or Alaska Native + Asian + Black or African American + Native Hawaiian or Other Pacific Islander + White + Two or more races	Equal to	Total non-Hispanic students
<b>Race/Ethnicity</b>	Native American or Alaska Native + Asian + Black or African American + Native Hawaiian or Other Pacific Islander + White + Two or more races + Hispanic	Equal to	Unduplicated count of students

### Compare Across Files

This involves comparing the values from one set of file specs with the values in another file spec and understanding when each set of values should be less than, equal to, or more than another set of values.

#### *Career and Academic Outcomes*

Each career/academic outcome while in facility (FS 218/219)<sup>1</sup> should be less than or equal to the total number of unduplicated students. There cannot be more students achieving an outcome than there are students in total.

Outcomes <sup>2</sup>	Equation	Total
Count of students who earned high school course credits (FS 218/219)	less than or equal to	Unduplicated count of students (FS 119/127)
Count of students who enrolled in GED program (FS 218/219)	less than or equal to	Unduplicated count of students (FS 119/127)

<sup>1</sup> Prior to SY 2022-23, the file spec is FS 180 for career/academic outcomes while in facility for both SA and LEA.

<sup>2</sup> It is okay if the total of all outcomes is more than the count of unduplicated students since students can achieve more than one outcome.

Count of students who obtained high school diploma or earned a GED (FS 218/219)	less than or equal to	Unduplicated count of students (FS 119/127)
Count of students who enrolled in job training courses/programs (FS 218/219)	less than or equal to	Unduplicated count of students (FS 119/127)
Count of students who obtained employment (FS 218/219)	less than or equal to	Unduplicated count of students (FS 119/127)
Count of students who were accepted and/or enrolled into postsecondary education (FS 218/219)	less than or equal to	Unduplicated count of students (FS 119/127)

### Academic Achievement

Users can also conduct the same type of data quality check on career/academic outcomes up to 90 days after exit (FS 220/221) to see if each outcome after exit is less than or equal to the unduplicated count of students.

Academic Achievement (FS 113/125)	Equation	Total (FS 119/127)
<ul style="list-style-type: none"> <li>Number of LT (Long-term) students who improved more than 1 grade from initial to follow up assessment</li> </ul> <p style="text-align: center;">+</p>	Less than or equal to	Total number of LT students
<ul style="list-style-type: none"> <li>Number of LT students who improved up to one 1 grade from initial to follow up assessment</li> </ul> <p style="text-align: center;">+</p>		
<ul style="list-style-type: none"> <li>Number of LT students whose assessment score did not change from initial to follow up assessment</li> </ul> <p style="text-align: center;">+</p>		
<ul style="list-style-type: none"> <li>Number of LT students whose assessment score decreased from initial to follow-up assessment</li> </ul>		

### Percentage Change

This involves calculating the percentage by which the number of students has increased or decreased. This is useful to identify any large jumps or dips in the student population. This calculation can be used to check changes in participation (FS 119/127), and changes in students achieving outcomes (FS 218/219).

#### Percentage change equation

$$\frac{(\text{Count of students in current year} - \text{Count of students in previous year})}{\text{Count of students in previous year}} \times 100$$

#### Example equation of percentage change of students with disabilities

$$\frac{(\text{Number of students with disabilities SY 2022-23} - \text{Number of students with disabilities SY 2021-22})}{\text{Number of students with disabilities SY 2021-22}} \times 100$$

*Example calculation of percentage of students with disabilities*

In SY 2020-21 the number of students with disabilities served by TIPD in the United States was 12,436. In SY 2019-20 the number of students with disabilities served by TIPD in the United States was 18,801.

$$\frac{(12,436 - 18,801)}{18,801} \times 100$$

$$\frac{-6365}{18,801} \times 100$$

$$-0.3385 \times 100 = -33.85\%$$

From SY 2019-20 to SY 2020-21, there was a 34% decrease in the number of students with disabilities served by TIPD in the U.S. If there is a large percentage change, the State should be prepared to explain why that change occurred.

*Note: a large percentage change does not necessarily indicate a data issue, but it helps identify areas to investigate and explain. States that have smaller programs may be more likely to see a larger percentage change. For this reason, it can be helpful to check the counts.*

### Completeness

This involves checking that data are reported for items that are expected. For example, if the data on neglected programs are reported for file spec 119 (student participation and demographics), then data on count of programs/facilities of neglected programs should be reported in the manual entry of CSPR.

IF	THEN
Count of programs reporting data (FS119/127) > 0	Count of programs/ facilities (manual entry) > 0
Count of programs reporting data (FS119/127) > 0	Average Length of Stay (manual entry) > 0
Count of programs reporting data (FS119/127) > 0	Average number of days served (manual entry) > 0

For example, if data are reported for students in Adult Corrections, then the count of facilities, average length of stay, and average number of days served should also be reported for Adult Corrections. If the data reported is zero or null, then that would indicate data are missing for that reporting category.