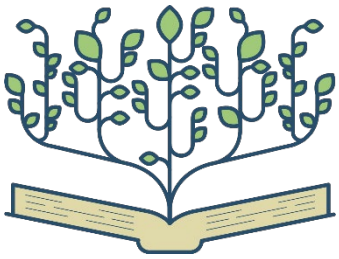

Section III: Data Quality

State coordinators can use the information in this section to:

- Identify common data quality challenges in Title I, Part D reporting
- Establish routines for data checks and internal validation
- Create clear documentation and training to support consistent reporting across grantees
- Write informative data notes to document data inconsistencies

Relevant appendix

- *Appendix D: Data Quality Calculations* – Provides step by step examples on how to set up and run calculations to detect potential data quality issues.

Data quality is the foundation for data-driven decision making. Data quality checks are a way to improve the accuracy of data. Data quality checks should be included throughout the data collection and reporting process. These checks can be included early on when facilities are collecting data by building it into data collection tools. Reviewing the submitted data using additional data quality checks will also reveal any potential data quality issues.

The business rules single inventory (BRSI) is a set of data quality rules that the U.S. Department of Education (ED) applies to ED*Facts* data submissions from State education agencies (SEAs). These rules are organized by file specification and fall into distinct types of data checks. There are four common ways to check data quality:

- **Completeness**: Checking that data are reported for all items that are required.
- **Subtotal to total**: Checking if the number of students in subgroups sums to the total number of students overall.
- **Expected relationship**: 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.
- **Year-to-year comparison**: Comparing the number reported in the current year to the previous year and identifying substantial differences. If the State has a small program, even minor changes in the actual count can result in large percentage changes from year to year. Keep this in mind when interpreting year-to-year comparisons.

For examples of each of these with calculations, please see Appendix D: Data Quality Calculations.

The data quality checklists below are organized by file specifications and are based on the BRSI. These rules are helpful to include throughout the data collection process. State coordinators can incorporate data quality checks into data collection forms, as well as conduct data quality checks when receiving the data, and prior to submitting. Coordinators can use the checklist when reviewing subgrantee data in addition to sharing the checklist with subgrantees to review their own data.

FS 119 (Subpart 1) and 127 (Subpart 2): Participation Checklist

For participation data, data quality considerations include verifying that the data are complete, confirming that demographic subgroups sum up to the total number of students, the number of long-term students is less than the total number of students, and ensuring the data are consistent from year to year. This is applicable to both Subpart 1 State agencies (SAs) and Subpart 2 local educational agencies (LEAs).

The desired response to each item in this table is "yes."

Question	YES	NO
Completeness: Does the data file contain at least one value that is not "0"? *Note: This check ensures the file is not made up entirely of zeros.*		
Completeness: Does the data file contain at least one value that is not "-1"? *Note: Since "-1" indicates a missing count, this check ensures the file is not entirely composed of missing values.*		
Completeness: Are the data complete for the data file (e.g., there are no blank fields or missing data)? *TIP: If there are no students in a subgroup (such as English learners), report 0s instead of leaving the fields blank.*		
Subtotal to total: Do the subtotals equal the totals?		
- Male + Female = Unduplicated count of students		
- Sum of ages (3 through 21) = Unduplicated count of students.		

- Native American/Alaska Native + Asian + Black/African American + Native Hawaiian/Pacific Islander + White + Two or more races + Hispanic = Unduplicated count of students		
Year-to-year comparison: Are there substantial changes in the count of students from the current year compared to the previous year?		
- (Student count previous year – Student count current year/Student count previous year) *100 >50% [Subpart 1 and Subpart 2]		
- (Student count previous year – student count current year)> 10 - [Subpart 1 only]		
- (Student count previous year – student count current year)> 50 [Subpart 2 only]		

FS 113 (Subpart 1) and 125 (Subpart 2): Academic Achievement Checklist

For academic achievement data, data quality considerations include verifying that the data are complete, confirming that the number of long-term students (students who are in facilities for at least 90 days) is equal to or exceeds the number of students who took an assessment, and ensuring that the data remain consistent from year to year. The goal is to ensure that **all** long-term students have received both an initial and follow-up assessment. Any section that applies only to Subpart 1 SAs or Subpart 2 LEAs is clearly indicated.

The desired response to each item in this table is "yes."

Question	YES	NO
Completeness: Does the data file contain at least one value that is not "0"? *Note: This check ensures the file is not made up entirely of zeros.*		
Completeness: Does the data file contain at least one value that is not "-1"? *Note: Since "-1" indicates a missing count, this check ensures the file is not entirely composed of missing values.*		
Completeness: Are the data complete for the data file (e.g., there are no blank fields or missing data)? *TIP: If there are no students in an achievement category (such as number who improved more than one grade level), report 0s instead of leaving the fields blank.*		
Expected relationship: Are there more students in total than assessment takers in total?		
- Total count of all students who participated in the program is greater than or equal to the count of students who took academic achievement assessments. **Subpart 1**: FS 119 (all participants) ≥ FS 113 (assessed students) **Subpart 2**: FS 127 (all participants) ≥ FS 125 (assessed students)		
Expected relationship: Are there more long-term students than reported assessment takers?		
- Total count of long-term students who participated in the program is greater than or equal to the count of students who took academic achievement assessments. **Subpart 1**: FS 119 (long-term participants) ≥ FS 113 (assessed students) **Subpart** 2: FS 127 (long-term participants) ≥ FS 125 (assessed students)		
Year to year comparison: Are there no substantial changes in the count of students from the current year compared to the previous year?		

Example: Percentage of students improving on assessments in previous year – Percentage of students improving on assessments in current year > 10		

FS 218 (Subpart 1) and 219 (Subpart 2): Outcomes In-Program Checklist

Outcome data should be reported for all students who participated in a Title I, Part D (TIPD) funded program. For Subpart 1 programs this includes both neglected and delinquent students. For Subpart 2 this includes both delinquent and at-risk students but does not include neglected students served with Title I, Part A funds. For outcomes in program data, data quality considerations include verifying that the data are complete and confirming that the number of students achieving an outcome does not exceed the total number of students served by the program. Any section that applies only to Subpart 1 SAs or Subpart 2 LEAs is clearly indicated. If a student achieves multiple outcomes (e.g., earns high school course credits and earns a high school diploma) the student should be reported in each category in which an outcome is achieved.

The desired response to each item in this table is "yes."

Question	YES	NO
Completeness: Does the data file contain at least one value that is not "0"? *Note: This check ensures the file is not made up entirely of zeros.*		
Completeness: Does the data file contain at least one value that is not "-1"? *Note: Since "-1" indicates a missing count, this check ensures the file is not entirely composed of missing values.*		
Completeness: Are the data complete for the data file (e.g., there are no blank fields or missing data)? *TIP: If there are no students who achieved an outcome (such as earned a GED), report 0s instead of leaving the fields blank.*		
Expected relationship: Is the number of students achieving outcomes less than or equal to the number of students reported?		
- Count of students who earned high school course credits is less than or equal to the unduplicated count of students who participated in the program. **Subpart 1**: FS 218 (course credits) ≤ FS 119 (participation) **Subpart 2**: FS 219 (course credits) ≤ FS 127 (participation)		
- Count of students who enrolled in a GED program is less than or equal to the unduplicated count of students who participated in the program. **Subpart 1**: FS 218 (GED enrollment) ≤ FS 119 (participation) **Subpart 2**: FS 219 (GED enrollment) ≤ FS 127 (participation)		
- Count of students who obtained a high school diploma or earned a GED is less than or equal to the unduplicated count of students who participated in the program. **Subpart 1**: FS 218 (diploma or GED earned) ≤ FS 119 (participation) **Subpart 2**: FS 219 (diploma or GED earned) ≤ FS 127 (participation)		
- Count of students who enrolled in job training courses or programs is less than or equal to the unduplicated count of students who participated in the program. **Subpart 1**: FS 218 (job training enrollment) ≤ FS 119 (participation) **Subpart 2**: FS 219 (job training enrollment) ≤ FS 127 (participation)		
- Count of students who obtained employment is less than or equal to the unduplicated count of students who participated in the program. **Subpart 1**: FS 218 (employment) ≤ FS 119 (participation)		

Subpart 2: FS 219 (employment) ≤ FS 127 (participation)		
- Count of students who were accepted and/or enrolled in postsecondary education is less than or equal to the unduplicated count of students who participated in the program. **Subpart 1**: FS 218 (postsecondary acceptance/enrollment) ≤ FS 119 (participation) **Subpart 2**: FS 219 (postsecondary acceptance/enrollment) ≤ FS 127 (participation)		
Expected relationship: Is the count of students who earned a GED less than or equal to the count of students who enrolled in a GED program? **Subpart 1**: FS 218 (GED earned) ≤ FS 218 (GED enrollment) **Subpart 2**: FS 219 (GED earned) ≤ FS 219 (GED enrollment)		
Expected relationship: Is the count of students who obtained a high school diploma less than or equal to the count of students who earned high school course credits? **Subpart 1**: FS 218 (diploma) ≤ FS 218 (course credits) **Subpart 2**: FS 219 (diploma) ≤ FS 219 (course credits)		

FS 220 (Subpart 1) and 221 (Subpart 2): Outcomes After Program Exit Checklist

For outcomes after program exit data, data quality considerations include verifying that the data are complete and confirming that the total number of students exceed the number of students achieving an outcome. Any section that applies only to Subpart 1 SAs or Subpart 2 LEAs is clearly indicated. If the SEA is unable to report outcomes after program exit due to State law, then the State coordinator must submit a data note that includes reference to the specific statute.

Note: students may leave the program but not leave the facility. For example, a student may graduate from the GED program but remain incarcerated in the facility.

The desired response to each item in this table is "yes."

Question	YES	NO
Completeness: Does the data file contain at least one value that is not "0"? *Note: This check ensures the file is not made up entirely of zeros.*		
Completeness: Does the data file contain at least one value that is not "-1"? *Note: Since "-1" indicates a missing count, this check ensures the file is not entirely composed of missing values.*		
Completeness: Are the data complete for the data file (e.g., there are no blank fields or missing data)? *TIP: If there are no students who achieved an outcome (such as earned a GED after exit), report 0s instead of leaving the fields blank.*		
Expected relationship: Is the count of students achieving outcomes less than or equal to total students reported?		
- Count of students who earned high school course credits after exit is less than or equal to the unduplicated count of students who participated in the program. **Subpart 1**: FS 220 (course credits) ≤ FS 119 (participation) **Subpart 2**: FS 221 (course credits) ≤ FS 127 (participation)		

- Count of students who enrolled in a GED program after exit is less than or equal to the unduplicated count of students who participated in the program. **Subpart 1**: FS 220 (GED enrollment) ≤ FS 119 (participation) **Subpart 2**: FS 221 (GED enrollment) ≤ FS 127 (participation)		
- Count of students who obtained a high school diploma or earned a GED after exit is less than or equal to the unduplicated count of students who participated in the program. **Subpart 1**: FS 220 (diploma or GED earned) ≤ FS 119 (participation) **Subpart 2**: FS 221 (diploma or GED earned) ≤ FS 127 (participation)		
- Count of students who enrolled in job training courses or programs after exit is less than or equal to the unduplicated count of students who participated in the program. **Subpart 1**: FS 220 (job training enrollment) ≤ FS 119 (participation) **Subpart 2**: FS 221 (job training enrollment) ≤ FS 127 (participation)		
- Count of students who obtained employment after exit is less than or equal to the unduplicated count of students who participated in the program. **Subpart 1**: FS 220 (employment) ≤ FS 119 (participation) **Subpart 2**: FS 221 (employment) ≤ FS 127 (participation)		
- Count of students who were accepted and/or enrolled into postsecondary education after exit is less than or equal to the unduplicated count of students who participated in the program. **Subpart 1**: FS 220 (postsecondary enrollment) ≤ FS 119 (participation) **Subpart 2**: FS 221 (postsecondary enrollment) ≤ FS 127 (participation)		
Expected relationship: Is the count of students who earned a GED less than or equal to the count of students who enrolled in a GED program? **Subpart 1**: FS 220 (GED earned) ≤ FS 220 (GED enrollment) **Subpart 2**: FS 221 (GED earned) ≤ FS 221 (GED enrollment)		
Expected relationship: Is the count of students who obtained a high school diploma less than or equal to the count of students who earned high school course credits? **Subpart 1**: FS 220 (diploma) ≤ FS 220 (course credits) **Subpart 2**: FS 221 (diploma) ≤ FS 221 (course credits)		

FS 224 (Subpart 1) and 225 (Subpart 2): Statewide Standardized Assessment Checklist

The desired response to each item in this table is "yes."

Question	YES	NO
Completeness: Does the data file contain at least one value that is not "0"? *Note: This check ensures the file is not made up entirely of zeros.*		
Completeness: Does the data file contain at least one value that is not "-1"? *Note: Since "-1" indicates a missing count, this check ensures the file is not entirely composed of missing values.*		

CSPR II Checklist

When reporting data for the Consolidated State Performance Report (CSPR) II, the State coordinator should ensure that the program/facility types reported in ED*Facts* match the program/facility types reported for the CSPR II. The average length of stay should be less than or equal to 365 days because the

reporting timeframe is limited to one year. Additionally, the average number of days served should be less than or equal to the average length of stay because students cannot be served more days than they stayed. In many cases, the number of days served is expected to be less than the length of stay since students may not be served every day (e.g. weekends, holidays).

The desired response to each item in this table is "yes."

Question	YES	NO
Expected relationship: Is the average length of stay less than or equal to 365 days?		
Expected relationship: Is the average number of days served less than or equal to the average length of stay?		
Expected relationship: Does the count of programs/facilities manually entered in CSPR match the corresponding program/facility data reported in ED*Facts* in FS 119 (participation, Subpart 1) and FS 127 (participation, Subpart 2)? *For example: If CSPR shows that one or more shelters are receiving funds (e.g., 2 shelters), then EDFacts should report a positive number of students in shelters. Reporting 0 students in EDFacts would indicate a discrepancy, since it conflicts with the CSPR data.*		
Expected relationship: If average length of stay > 0 is this also reflected in the count of programs reporting data? *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.*		
Expected relationship: If the average count of programs reporting data in FS 119 (participation, Subpart 1) or FS 127 (participation, Subpart 2) > 0 then is the average number of days served > 0?		

Troubleshooting Data Issues

When a data issue is identified, there are steps to investigate and address the data issue:

1. Identify the data discrepancy
2. Investigate the potential data entry or reporting issue
3. Identify potential causes of data discrepancy
4. Add a data note to explain the data discrepancy or update data entry to correct the value

If there is a potential data error, it is important to investigate the issue and understand what is causing the discrepancy. The data flag could be the result of a data error, or it could be reflective of external changes.

For example, the number of students earning high school course credits may be higher than the total number of students. Once such an issue is identified, it is important to contact the facility to understand how the data were reported. In this case, it could be a data entry error, such as reporting the total number of credits earned instead of the number of students who earned a credit, or a number was mistakenly entered. Working with the facility to understand the discrepancy is critical to update and submit accurate data.

It is advisable to conduct data checks on two occasions. The first review should be conducted when each program/SA/LEA submits data, allowing time to request corrections if needed. A second review should take place upon submission to ensure that data was entered into the system correctly.

Data Notes

When submitting data, it is important to include data notes to explain the reasons behind any anomalies, unexpected changes, or data limitations. Data notes demonstrate that the individual submitting the data understands what is causing the data flag—whether it is a change in program participation, a reporting limitation, or an external factor like a policy shift. This context is essential for accurate interpretation and helps reviewers distinguish between true data quality problems and explainable fluctuations.

Well-written data notes can:

- Provide transparency by explaining why significant changes occurred (e.g., a program closed due to low enrollment and small allocations, leading to a large drop in student counts).
- Demonstrate ownership and accountability (e.g., an SEA acknowledging incomplete demographic data and committing to improve reporting with the LEA).
- Clarify operational constraints that affect data collection (e.g., students leaving the program before follow-up assessments could be administered).
- Build trust with reviewers by showing that the data are being actively monitored, understood, and improved where needed.

Including clear, specific data notes ensures that data users can make well-informed decisions and assessments based on a complete understanding of the data. Below are some examples of data notes:

File Spec	BRSI Rule Failure	Data Note Example
113	**Year to Year Comparison**: The SEA-level data indicate a change of -60% between the current school year number of long-term students who showed improvement of up to one full grade level or more on reading/language arts initial and follow-up assessments and the prior school year number of long-term students who showed improvement of up to one full grade level or more on reading/language arts initial and follow-up assessments for the following TIPD, Subpart 1 Programs types: Adult correction.	The significant decrease is due to the very small number of participating students in the TIPD, Subpart 1 program. There was a decrease in number of long-term students from 2023-24 to long-term students participating in 2024-2025. This led to a significant percentage change. The only eligible State agency (Department of Corrections) decided to discontinue operating their TIPD, Subpart 1 program, due to the small number of eligible students and the resulting small allocation which required great administrative needs.
127	**Category Set Comparison**: In one LEA, the sum of the Category Set A [N or D Program (Subpart 2), Racial Ethnic] does not equal Subtotal 1 [N or D Program (Subpart 2)] for the following TIPD, Subpart 2 program: Juvenile correction. These two counts should equal.	The data are accurate as reported. The LEA reported racial/ethnic data for some students, but indicated the data were not available for all the students. The SEA will work with the LEA to ensure complete and accurate data can be reported moving forward.
125	**Completeness**: In 1 LEA(s), no data or all zeros were reported for the sum of students with initial and follow-up assessment results in reading/language arts for the following TIPD, Subpart 2 program: at-risk programs.	This program was not able to collect the follow-up assessment results. Students are often pulled out of the program with little or no warning and do not get the opportunity to take the follow-up assessment before they leave.

In the first example, the data note clearly explains the changes occurring in the facilities which affected the number of long-term students. In the second example, the data note explains the reason for the data discrepancy. The third and fourth examples highlight data discrepancies that relate to data collection challenges. For more tips on writing data notes please see this Department of Education resource.

Ensuring data quality is a critical step in the data collection and reporting process that enables States and stakeholders to use the data to understand and improve programs for students served by TIPD.

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
Gender	Male + Female	Equal to	Unduplicated count of students
Ethnicity	Non-Hispanic + Hispanic	Equal to	Unduplicated count of students
Age	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
Race	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
Race/Ethnicity	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)[1] 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[2]	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)

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

[2] 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)
• Number of LT (Long-term) students who improved more than 1 grade from initial to follow up assessment +	Less than or equal to	Total number of LT students
• Number of LT students who improved up to one 1 grade from initial to follow up assessment +		
• Number of LT students whose assessment score did not change from initial to follow up assessment +		
• Number of LT students whose assessment score decreased from initial to follow-up assessment		

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

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

Example equation of percentage change of students with disabilities

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

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.