MEASUREMENT INCORPORATED

[updated December 29, 2022]

EZReports Data Verification
Guidelines for RC Data Specialists and Local Data Managers

NOTES:

The purpose of this document is to provide guidelines for one set of methods that can be used to meet several of the key objectives for verifying EZReports data; however, these methods do not represent the only way to accomplish each objective. While they provide an efficient approach to reviewing large numbers of cases, they require a degree of familiarity with Microsoft Excel.

Using Excel has the advantage that the data reports generated by EZReports are downloaded into Excel files, and data uploads to EZReports must also be in Excel; however, Excel files can be easily converted or imported directly into many other data management and analysis applications.

IMPORTANT:

! In all of the following data manipulations, ***do not make any changes to the original data fields downloaded from EZReports.***

! When copying formulas to the spreadsheet, ***copy the formula exactly*** (for example, it is important to use the " character and not the “ or ” characters in the formulas); then edit as needed (for example to replace ***CR*** with correct cell addresses).

! If the Excel syntax does not work, first ***make sure it is preceded by an equals sign***, and make sure the cell format = General. You may need to double click the cell and close it again after changing the format.

! Once you have confirmed that the formulas are working correctly, you need to save values in order for functions such as Find, Pivot Tables, Sorts, and numeric calculations to work properly.

- ***Keep a copy of the formula*** (with the equals sign omitted) in the cell above the first row of data so that you have a record of how the value was determined.
- In the data rows, ***convert the formulas to values*** by using Copy – Paste Special – Values and number formats.

! All of the steps described below to check whether data points are “valid” only indicate whether they are in the correct format. ***They do not tell you whether they are accurate.*** (*e.g.* a valid Student ID may belong to a different student, and a valid School Code may reference the wrong school.)

! Accuracy of School Codes can be checked as described under “Verify alignment of school names” in the last section, but accuracy of student IDs is much more difficult to check. For this reason, ***it is critical that programs ensure that they are entering accurate student ID codes into EZReports***. Inaccurate student IDs will likely result in those students being dropped from analyses, potentially resulting in misleading findings about program activities and outcomes.

MEASUREMENT INCORPORATED

[updated December 29, 2022]

Download a student demographics report using EZR Report Wizard

Purpose: This database will include all records and fields needed to perform the validation steps addressed in this document. These include records for all students who had any program participation during the program year.

- **Report Type:** Generate Student List of Demographics
- **Report Input:**
- ***Student Status*** = All;
- Filter for ***Location***, *or* generate for entire site or program. (RCs can filter for Region after data are downloaded.)
- Leave all other fields blank to ensure that all registered students are listed.
- **Report Fields:**

  Select all desired fields; to include, *at least*, ***Date of Birth, EZReports ID, NYSSIS State StudentID, District Student ID, Student School, Grade Level***.

The file will download to an Excel 97-2003 Workbook (*.xls) file; re-save the file as an Excel 2010 Workbook (*.xlsx), which will create a much smaller file.

Ensure that all participants have valid State ID codes, and OSIS ID codes for NYC students

Purpose: Valid State IDs (NYSSIS IDs) always have 10 digits, no non-numeric characters, and never begin with leading 0s. Valid OSIS IDs always have 9 digits, no non-numeric characters, and never begin with leading 0s. The process also locates and highlights placeholders in the ID fields. These placeholders, as defined by NYSED, can only take one of two forms, as detailed below:

State ID field: 008nnnxxxx

District ID field: 08nnnxxxx

All placeholders, missing records and invalid entries require follow-up, and must be corrected before finalizing data for submitting to SIRS for APRs and statewide analyses, for documentation of participation targets, or for submitting to participant schools/districts to obtain data for reporting on local objectives.

1. **Identify invalid ID codes from above report:**

   a. ***NYSSIS IDs*** are 10 characters, all numeric, none beginning with 0.

   The established format for a **placeholder** code in the ***NYSSIS ID*** field is ***008nnnxxxx***, where 8nnn is the 4-digit grant number, and xxxx is any 4-digit number used by the program to uniquely identify each student. Note however that the Excel file may drop leading 0s. As a result, a valid place holder code must be one of the following:

   - 10 digits beginning with 008, or
   - 8 digits beginning with 8.

1. Identify invalid ID codes (continued):

Create a flag indicating whether each code in the NYSSIS ID field is valid:

- Add a new column labeled **Valid NYSSIS ID**?, with column Format = General, with the following syntax in each row:

  =IF(ISBLANK(CR),"Missing",IF(AND(VALUE(RIGHT(CR,10))>=1000000000,VALUE(RIGHT(CR,10))<=9999999999),"Valid",IF(AND(VALUE(RIGHT(CR,10))>=80000000,VALUE(RIGHT(CR,10))<=89999999),"Placeholder","Invalid")))

  …where ***CR*** = ***C***olumn and ***R***ow of the cell containing the student's NYSSIS ID.

  Make sure to copy the correct numbers of 0s and 9s in the above formula. Applied correctly, in most cases this will return values of either ***Valid***, ***Placeholder***, ***Invalid*** or ***Missing***. However, this depends on EZReports using a validation that only allows all-numeric entries. If for any reason there are any non-numeric characters, this syntax will return a value of ***#VALUE!***, which should be interpreted the same as ***Invalid***.

b. [For NYC schools only]: ***District IDs*** are 9 characters, all numeric, none beginning with 0.

The established format for a **placeholder** code in the ***District ID*** field is ***08nnnxxxx***, where 8nnn is the 4-digit grant number, and xxxx is any 4-digit number used by the program to uniquely identify each student. Note however that the Excel file may drop leading 0s. As a result, a valid place holder code must be one of the following:

- 9 digits beginning with 08, or
- 8 digits beginning with 8.

Create a flag indicating whether each code in the *District ID* field is valid:

- Add new column, **Valid OSIS ID?**, with column Format = General, with the following syntax in each row:

  =IF(ISBLANK(***CR***),"Missing",IF(AND(VALUE(RIGHT(***CR***,9))>=100000000,VALUE(RIGHT(***CR***,9))<=999999999),"Valid",IF(AND(VALUE(RIGHT(***CR***,9))>=80000000,VALUE(RIGHT(***CR***,9))<=89999999),"Placeholder","Invalid")))

  …where ***CR*** = ***C***olumn and ***R***ow of the cell containing the student's ***District ID***.

  Make sure to copy the correct numbers of 0s and 9s. This will return values of either ***Valid***, ***Placeholder***, ***Invalid*** or ***Missing***. However, if there are any non-numeric characters, this syntax will return a value of ***#VALUE!***, which should be interpreted the same as ***Invalid***.

2. Identify distribution of invalid ID codes by program, by site, and/or by grade level

a. If you wish to aggregate data by Grant, create a new column, **Grant #**, that extracts the 4-digit grant # from the Site Name using:

RIGHT(TEXTBEFORE(***CR***," - "),4)

…where ***CR*** = ***C***olumn and ***R***ow of the cell containing the student's ***Name of Site***.

Note that this formula will not always work, for example if there is more than one hyphen in the Site Name, or if the Site Name does not contain the grant #.

b. Create a pivot table* summarizing:

   i. # and % of students where **NYSSIS ID Valid?** = ***Placeholder*** at desired aggregation level,

   ii. # and % of students where **NYSSIS ID Valid?** = ***Invalid, #VALUE!*** or ***Missing*** at desired aggregation level,

   iii. [in NYC programs] # and % of students where **District ID Valid?** = ***invalid, #VALUE!*** or ***Missing*** at desired aggregation level.

* Refer to the following links if you need help using Pivot Tables:

https://www.youtube.com/watch?v=qu-AK0Hv0b4 [very brief introduction, 6 min.]

https://www.youtube.com/watch?v=h4XOLbO20Ig [full course for beginners, 1 hr. 20 min.]

[Note: I have not vetted these videos for quality.]

c. Use Conditional Formatting to highlight programs/sites/grades with largest #/% of problems, correct any ***invalid***, ***#VALUE!, placeholder*** or ***Missing*** codes.

Confirm duplicate student ID codes are resolved

Purpose: Duplicate ID codes may occur for several reasons:

- *Invalid duplicates. Because of record keeping or data entry errors, the same ID was entered accidentally for two (or more) different students.*
- *Valid duplicates. The same student has more than one record, e.g. because they moved to a different program site during the program year, or because their data was entered separately during summer activities vs. school year activities.*

Invalid duplicates require correcting the ID codes. For valid duplicates, all student demographics (name, DOB, grade, etc.) should be the same, but attendance records would be different. Hours should be merged in order to accurately report program participation targets and for reporting on local objectives.

1. **Identify duplicate ID codes from above report**

   a. Flag duplicate ID codes in ***District IDs*** and ***NYSSIS IDs*** using Conditional Formatting → Highlight Cell Rules → Duplicate Values

2. **Sort by ID and determine reason for duplicate values**

3. **Correct duplicates**

Confirm all other demographic variables are in correct format

Purpose: For all types of data summary, analysis and merging, all data must be in the correct format. Many variables in EZReports prevent invalid formats, but not all do; and very few ensure that data are within valid ranges of values.

- ***DOB*** = ***mm/dd/yyyy***, between (approximately) 06/30/2003 and 09/01/2018
- ***NYSSIS IDs*** and ***OSIS IDs*** (see above)
- ***School Code*** = 12 digits, all numerals, between ***010100010001*** and ***680801659999***, with last four digits between 0001-9999

1. **Identify invalid School Codes from above report:**

   Create a flag indicating whether each code in the School Code field is a valid BEDS code:

   - Add new column, **Valid School Code?**, with column Format = General, with the following syntax in each row:

     IF(ISBLANK(CR),"Missing",IF(OR(AND(LEN(CR)=11,VALUE(RIGHT(CR,11))>=10100010001,VALUE(RIGHT(CR,11))<=99000000000,VALUE(RIGHT(CR,4))>=1,VALUE(RIGHT(CR,4))<=9999),AND(LEN(CR)=12,VALUE(RIGHT(CR,12))>=10100010001,VALUE(RIGHT(CR,12))<=680801659999,VALUE(RIGHT(CR,4))>=1,VALUE(RIGHT(CR,4))<=9999)),"Valid","Invalid"))

     …where ***CR*** = ***C***olumn and ***R***ow of the cell containing the student's Day School BEDS code.

     This will return values of either ***Valid, Invalid*** or ***Missing***. However, if there are any non-numeric characters (*e.g.* in NYC where school codes are represented as PS999Q), this syntax will return a value of ***#VALUE!***, which should be interpreted the same as ***Invalid***.

2. **Identify distribution of invalid or missing School Codes by program, by site, and/or by grade level** using a pivot table summarizing # and % of students where **School Code Valid?** = ***Invalid, Missing*** *or* ***#VALUE!*** at desired aggregation level. (See Step 2a under validating Student ID codes to identify program/grant number.)

   Correct any incorrect format

Verify alignment of school/district names and BEDS codes with State records

Purpose: Analogous to student ID codes, school (and district) BEDS codes may be inaccurate due to record keeping or data entry errors. Unlike student IDs however, BEDS codes can be corrected fairly easily using public State records.

1. Download IMF Institutions report by County (https://eservices.nysed.gov/sedreports/list?id=1)
2. Compare ***School Name*** in EZReports to ***Popular Name*** or ***Legal Name*** in IMF report, by matching on BEDS codes (*i.e.* EZR ***School Code*** vs. IMF ***SED Code***) – make sure they identify the same school.

   If they do not identify the same school, or they identify the wrong school, you can reverse the above process: find the correct BEDS code by matching on School Name.

   Refer to the following link if you need help using VLookup:

   https://support.microsoft.com/en-us/office/vlookup-function-0bbc8083-26fe-4963-8ab8-93a18ad188a1
3. Correct inaccurate school codes.
4. You can perform the same check to confirm accuracy of District BEDS codes. If you have only School BEDS codes in your EZReports download, you can determine the BEDS code for the District from the School BEDS code:

   a. For all public schools other than charter schools, simply replace the last 4 digits of the School code with '0000' to get the District code, using the following syntax:

      =CONCATENATE(LEFT(***CR***,8),"0000")

      …where ***CR*** = ***C***olumn and ***R***ow of the cell containing the School BEDS code.

   b. For charter schools, the district code is the same as the school code. You can identify charter schools by searching for the word "Charter" in the school name, but a more reliable method is to use the BEDS code: all charter schools have the digits 86 in positions 7-8 of the 12-digit BEDS code.