The importance of data integrity and transparency to the evaluation system can't be underestimated. Data integrity requires verification and cleaning of data as well as clear procedures for data collection.

Reviewing Data Quality

Strategies should be in place to review data quality during and after data collection. Examples include:

- looking at the first wave of responses and number of 'no response' or refusals
- keeping data collectors and the evaluation lead connected
- after data entry, double checking for quality and consistency by sorting to find missing, high or low values (quantitative), or by reviewing transcripts entered (qualitative)

Establish a Data Validation Process

When establishing a process, consider the following questions:

- What validation process can be established to ensure clean data? (e.g., administrators monitoring input)
- Have criteria been established to ensure data confidentiality?
- What training will data collectors need to ensure accurate data collection?
- Which staff will require training to ensure accuracy in data entry and reporting?

Cleaning Data

Checking data for errors is commonly called "cleaning." Cleaning data is critical because "dirty" data can severely influence results. Cleaning data can also be time-consuming, so be sure to plan accordingly.

Three most commonly used in cleaning methods are:

- Spot-checking
- Eye-balling
- Logic checks

The best practice is to use all three approaches in order to catch all possible errors. The following is table that summarizes each approach.

Spot-checking	Eye-balling	Logic checks
This technique involves comparing the raw data to the electronically entered data to check for data-entry and coding errors. To spot-check quantitative survey data, randomly select several participants' completed paper surveys and compare them to the data on the electronic spreadsheet. For qualitative data check whether participants' words were transcribed accurately and are attributed to the right individual. If no errors are found in the first round of spot-checking, randomly check another round of the raw data. If multiple errors are found, and it is clear that it was not an isolated incident, all of the raw data will need to be checked to ensure that each record was entered correctly.	This technique involves reviewing the data for errors that may have resulted from a data-entry or coding mistake. Example: if a survey question reads: "Did you participate in the summer program?", participants can only respond with a "no" or "yes." This question is coded for data entry so that "No" is assigned a value of 0, while "yes" responses are assigned a value of 1. Therefore, any number other than a 0 or 1 in the response column for this spreadsheet would be an obvious error. If these errors are found, the original raw data will need to be found and the participant's answer entered correctly.	This technique involves a careful review of the electronically entered data to make sure that the answers to the different questions "make sense." Example, if participant A indicated they did not attend the summer program then it would be illogical for this participant to have provided a satisfaction rating to a follow-up question asking then to rate their experience at the program. As with the other types of errors, it will be necessary to go back to the original raw data for that participant and enter the correct data instead.

Source used:

American Institute for Research: Centre on Great Teachers and Leaders
https://gtlcenter.org/tools-publications/online-tools/principal-evaluation/component-6

Public Health Ontario
https://www.publichealthontario.ca/en/eRepository/At_A_Glance_Evaluation_2015.pdf

The Pell Institute and Pathways to College Networks: Evaluation Toolkit
http://toolkit.pellinstitute.org/evaluation-guide/analyze/enter-organize-clean-data/