# Evaluating Survey Data Quality

The Total Survey Error (TSE) paradigm is a framework widely used to assess and enhance the quality of survey data. It addresses all potential sources of error throughout the survey process, from design and data collection to processing and analysis. The following factors for evaluating survey data quality are adapted from the TSE paradigm (see references).
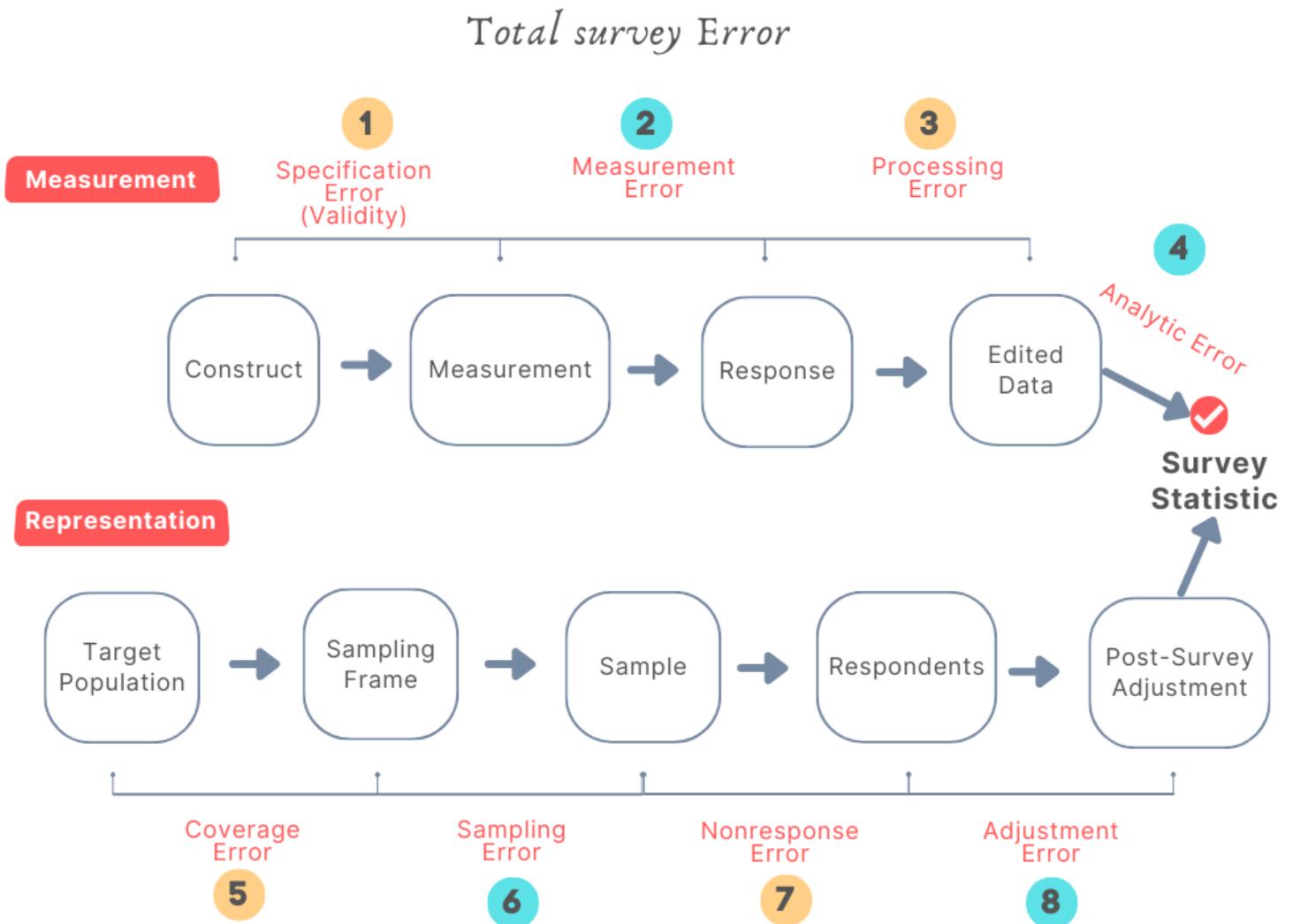
Measurement Aspect

1. **Specification Error (Validity):** Do the data variables accurately measure the theoretical construct they are intended to measure? For example, are variables such as household income, education, or household wealth valid indicators to measure socioeconomic status?

2. **Measurement Error:** Does the survey instrument accurately measure what it is intended to measure?
   a. **Respondents** may deliberately or unintentionally provide incorrect information. For example, respondents agree or disagree with statements regardless of their content (response style behaviors); use less effort to provide optimal responses (satisficing); are unable to remember information accurately (recall bias); answer questions in a manner that will be viewed favorably by others (social desirability bias); or withhold accurate information on sensitive issues like drug use or sexual behavior.
   b. **Interviewers** may inappropriately influence responses or record the responses incorrectly.
   c. **Questionnaires** may be designed with unclear terms or jargon, ambiguous or leading questions, confusing instructions, or inadequate response options.
   d. **Mode of Data Collection:** There may be differences in responses due to the data collection method (e.g., online, phone, face-to-face) (also called "mode effect").

3. **Processing Error:** Does the edited data accurately capture the survey responses? Are there any errors caused by data entry, coding, outlier editing, assignment of survey weights, or non-response imputing?
4. **Analytic Errors** arise in the post-processing steps after data has been collected from the field and stored in an analytic dataset. Errors may arise from incorrect merging, attribution of response to the wrong individual, incorrect use of survey weights, design features for estimation and inference, etc.

Representation Aspect

5. **Coverage Error:** Does the sampling frame represent the target population? Are some members of the target population excluded from the sampling frame? Examples include individuals without phone access in phone surveys, those without a permanent address in address-based sampling, or institutionalized populations (e.g., prisoners or dormitory residents), who often face undercoverage. The sampling frame may also contain errors such as omissions, duplicates, incorrect inclusions, and non-up-to-date information.
6. **Sampling Error:** Is the sample selected representative of the target population?
   a. **Sample Scheme:** Is the sample selected using probability-based random sampling methods, or less reliable non-probability methods such as online polls, opt-in panels, convenience sampling, or interactive voice response methods (e.g., robocalls or automated calls)?
   b. **Sample Size:** Is the sample size sufficient to ensure reliable and accurate results?
   c. **Estimator:** Are the statistical methods, formulas, or algorithms used to estimate population parameters robust and unbiased?
7. **Nonresponse Error:** Are there gaps between respondents and the sample?
   a. **Unit nonresponse** occurs when a sample unit (individual, household, or organization) does not respond to any part of the questionnaire;
   b. **Item nonresponse** occurs when the questionnaire is only partially completed and some items are not answered;
   c. **Incomplete response** occurs when the response to an open-ended question is incomplete or very short and inadequate;
   d. **Panel attrition** occurs when a sample unit is lost over the period of a longitudinal study.
8. **Adjustment Error:** Does the adjustment or correction made after survey data is collected reduce its accuracy or introduce new errors? Adjustment errors can happen in the following instances:
   a. **Weighting Adjustments** may be applied to correct for unequal selection probabilities, nonresponse, or to align with known population characteristics;
   b. **Post-Survey Adjustments** such as imputing missing values or making corrections based on identified biases;
   c. **Nonresponse Adjustments** such as methods used to address and account for nonrespondents in the sample.

Total Survey Quality

The TSE focuses on the accuracy dimension of survey quality. Total survey quality is also dependent on other non-statistical quality dimensions:

- **Credibility:** Is the data collection methodology credible and considered trustworthy by the survey community?
- **Comparability:** Are demographic, spatial, and temporal comparisons valid?
- **Usability/Interpretability:** Is the documentation clear, and is the metadata well managed?
- **Relevance:** Does the data satisfy the researcher's needs?
- **Accessibility:** Is access to the data user-friendly?
- **Timeliness and Punctuality:** Does data delivery adhere to the schedule?
- **Completeness:** Are the data rich enough to meet analysis objectives without placing undue burden on respondents?
- **Coherence:** Can estimates from different sources be reliably combined?

For details on assessing survey data's fitness for use and engaging in discussions with data providers and researchers, refer to **Data Quality Literacy Series 09 Commercial Data Quality: Conversation with the Vendors and 10 Commercial Data Quality: Conversation with the Researchers**.

To learn more about survey data, refer to **Data Quality Literacy Series 12: Understanding Survey Data and Public Poll**.

References

Biemer, P. P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly, 74*(5), 817-848. https://academic.oup.com/poq/article/74/5/817/1815551

Groves, R.M., and Lyberg, L. (2010). Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly, 74*(5), 849–879.

ITSEW, Lyberg, L. & Inizio. (2019). *Total Survey Error: Roots and Evolution.* INIZIO Presentation. https://www.niss.org/sites/default/files/ITSEW2019%20Primer%20-%20Lyberg.pdf

Liu, G., Bordelon, B., Nagar, R., Sarti, J., Nguyen, U., & Boettcher, J. (2024). *Data Quality Literacy: A Guidebook.* Institute of Museum and Library Services (IMLS) Grant Project. https://doi.org/10.31219/osf.io/ruawm

Maslovskaya, O. (n.d.). *Data Quality: Total Survey Error (TSE).* National Center for Research Methods Report. https://www.ncrm.ac.uk/resources/online/data_quality_and_survey_error/downloads/slides/data_quality_total_survey_error.pdf

Moya, C. (n.d.). *Inferential Statistics and Complex Surveys: Chapter 5 Total Survey Error Framework.* https://bookdown.org/cristobalmoya/iscs_materials/tse.html

West, B. & Schulz, P. (2018). *Total Survey Error: A Framework for High-quality Survey Design.* https://pdhp.isr.umich.edu/workshops/total-survey-error-a-framework-for-high-quality-survey-design-with-west-and-schulz/

Visit the project website to learn more!
https://www.dataqualityliteracy.org

INSTITUTE of
Museum and **Library**
SERVICES