Data Collection Methods
An introduction

Introduction

- Data collection means gathering information to address those critical evaluation/ research questions that you have identified earlier in the evaluation/ research process.

- To plan data collection, you must think about the questions to be answered and the information sources available.

Cont,

- You must begin to think ahead about how the information could be organized, analyzed, interpreted and then reported to various audiences

- There are many methods available to gather information, and a wide variety of information sources.

Pre-Data Collection Steps

1. Clearly define the goals and objectives of the data collection
2. Reach understanding and agreement on operational definitions and methodology for the data collection plan
3. Ensure data collection (and measurement) repeatability, reproducibility, accuracy, and stability

What kind of data should be collected?

- The information you collect is the evidence you will have available to answer the evaluation/research questions.
- Poor evidence is information which cannot be trusted, is limited, or simply is not relevant to the questions asked.
- Good evidence is information that comes from reliable sources and through trustworthy methods that address important questions

Where do data come from?

- Take a step back – if we're starting from scratch, how do we collect / find data?
  - Secondary data
  - Primary data

Sources of Data

- Primary sources
  - Primary data refer to information obtained firsthand by the researcher on the variables of interest for the specific purposes of the study
- Secondary sources
  - Secondary data refer to information gathered from sources already existing
  - data someone else has collected

Secondary Data – Examples of Sources

- County health departments
- Vital Statistics – birth, death certificates
- Hospital, clinic, school nurse records
- Private and foundation databases
- City and county governments
- Surveillance data from state government programs
- National agency statistics – Census, etc.

Secondary Data – Limitations

- Let us look at statistics in the https://www.knbs.or.ke/
- Any accuracy on this data?Why would you accredit this source?
- What were the resources needed for the success of such data?

Secondary Data – Limitations

- When was it collected? For how long?
  - May be out of date for what you want to analyze.
  - May not have been collected long enough for detecting trends.

Secondary Data – Limitations

- Is the data set complete?
  - There may be missing information on some observations
  - Unless such missing information is caught and corrected for, analysis will be biased.

Secondary Data – Limitations

- Are there confounding problems?
  - Sample selection bias?
  - Source choice bias?
  - In time series, did some observations drop out over time?

Secondary Data – Limitations

- Are the data consistent/reliable?
  - Did variables drop out over time?
  - Did variables change in definition over time?
    - E.g. number of years of education versus highest degree obtained.

Secondary Data – Limitations

- Is the information exactly what you need?
  - In some cases, may have to use "proxy variables" – variables that may approximate something you really wanted to measure. Are they reliable? Is there correlation to what you actually want to measure?

Secondary Data – Advantages

- No need to reinvent the wheel.
  - If someone has already found the data, take advantage of it.

Secondary Data – Advantages

- It will save you money.
  - Even if you have to pay for access, often it is cheaper in terms of money than collecting your own data. (more on this later.)

Secondary Data – Advantages

- It will save you time.
  - Primary data collection is very time consuming.

Secondary Data – Advantages

- It may be very accurate.
  - When especially a government agency has collected the data, incredible amounts of time and money went into it. It's probably highly accurate.

Secondary Data – Advantages

- It has great exploratory value
  - Exploring research questions and formulating hypothesis to test.

Primary Data

- Primary data – data you collect

Primary Sources of Data

- Individuals
- Focus groups
  - Aimed at obtaining respondents' impressions, interpretations, and opinions.
  - Provides only qualitative and not quantitative information
  - Can not be considered to be truly representative
  - Focus groups are used for (1) exploratory studies, (2) making generalizations based on the information gathered by them, and (3) conducting sample surveys

Primary sources

- Panels
  - Whereas focus groups meet for a one-time group session, panels meet more than once.
  - Static or dynamic
  - Typically used when several aspects of a product are to be studied from time to time
- Unobtrusive Measures
  - Originate from a primary source that does not involve people

Primary Data - Limitations

- Do you have the time and money for:
  - Designing your collection instrument?
  - Selecting your population or sample?
  - Pretesting/piloting the instrument to work out sources of bias?
  - Administration of the instrument?
  - Entry/collation of data?

Primary Data - Limitations

- Uniqueness
  - May not be able to compare to other populations

Primary Data - Limitations

- Researcher error
  - Sample bias
  - Other confounding (is something, other than the thing being studied, that could be causing the results seen in a study) factors

Data collection choice

- What you must ask yourself:
  - Will the data answer my research question?

Data collection choice

- To answer that
  - You much first decide what your research question is
  - Then you need to decide what data/variables are needed to scientifically answer the question

Data collection choice

- If that data exist in secondary form, then use them to the extent you can, keeping in mind limitations.
- But if it does not, and you are able to fund primary collection, then it is the method of choice.

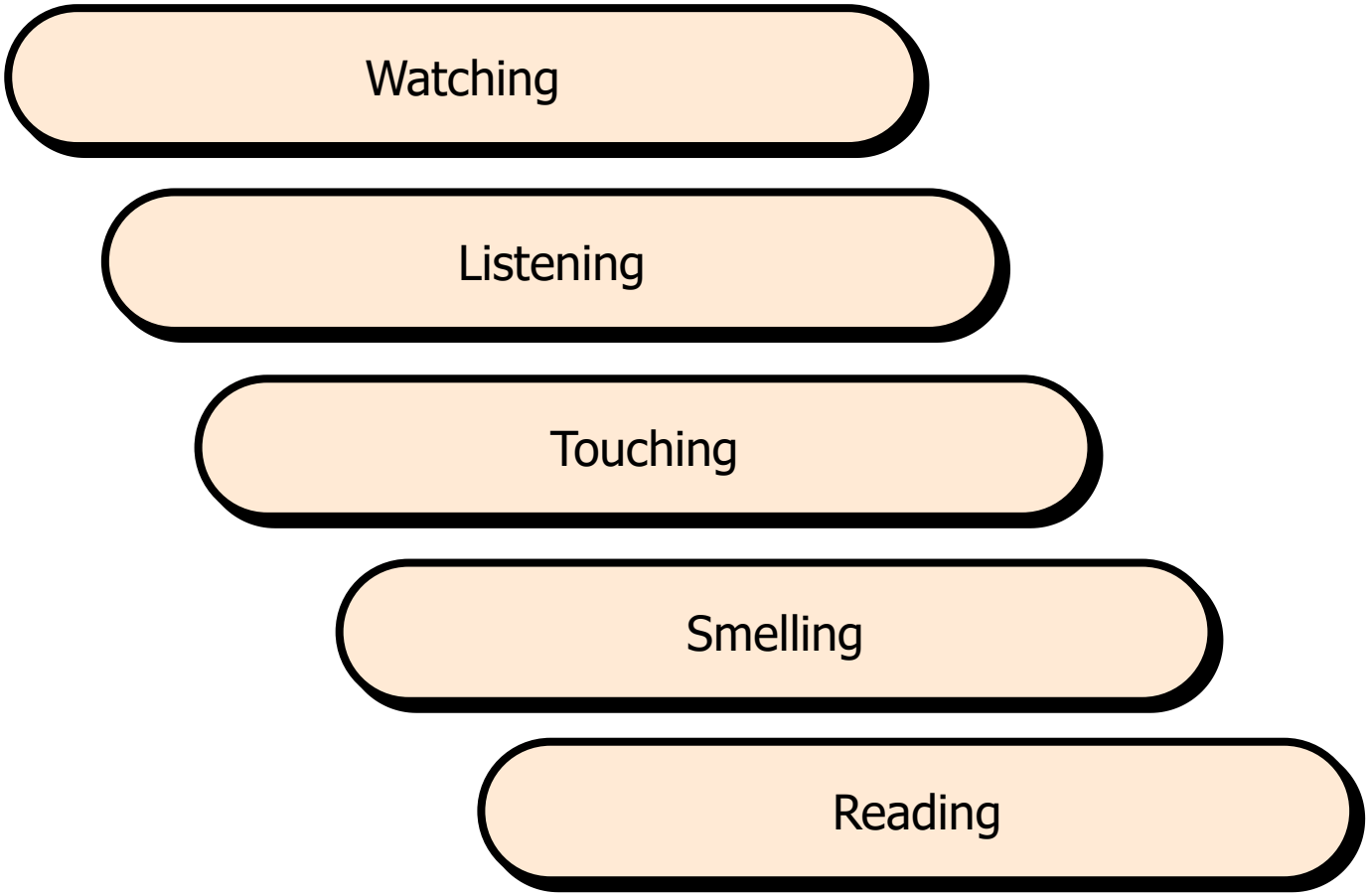
Primary Data - Examples

- Surveys
- Focus Group Discussions (FGDs)
- Experiments and Observational studies
- Questionnaires
- Interviews
- etc

Observation

- Nonparticipant and participant observer
- Structured and unstructured

Observation



Evaluation of Behavioral Observation

Strengths

- Securing information that is otherwise unavailable
- Avoiding participant filtering/ forgetting
- Securing environmental context
- Optimizing naturalness

Weaknesses

- Enduring long periods
- Incurring higher expenses
- Having lower reliability of inferences
- Quantifying data
- Keeping large records