Integrating evidence to tell the evaluation story

Greg Mason
Presentation to Canadian Institutes of Health Research
November 8, 9, 2012

Day 1- Afternoon
Adding characters

Goals of the workshop

- To present an intermediate and critical review of the main qualitative and quantitative lines of evidence currently used in evaluations
- To align these lines of evidence with the nature of the questions posed
- To review how to draw the lines of evidence into a coherent evaluation “story”

Outline of the workshop

Day 1 Morning – Creating the evaluation plot

Day 1 Afternoon – Adding characters

Day 2 Morning – Editing, assessing, interpreting

Day 2 Afternoon –Telling the story

Data, statistics, information – the characters

- **Data** – information at a unit record level
  - The responses of a single key informant
  - Attributes of a grant awarded to a specific researcher
  - Responses of a single respondent on a survey
  - *Simple data* … a single fact (e.g., age, opinion) of the unit record in response to a single question
  - *Complex data* … the collection of facts from the unit record
- **Statistic** – mathematical aggregation of data
  - Means, variances, sums, range, regressions, etc.
- **Information** – interpretation of statistics (comparisons, trends, projections, forecasts, implications, etc.)

Sources of data

- **Quantitative data** are countable (numbers - integers and real)

Example - sources

Administrative files (applications for funding, financial flows, client interactions, outcomes…)
Sample surveys using fixed response categories
Secondary analysis (official and public statistics)

- **Qualitative data** defy counting unless imposed

Example - sources

Pictures, music…
Text (verbatim interviews, self-reports, diaries, ethnographic descriptions, case studies…)

Collecting/Processing Quantitative Data (1) Administrative files

Advantages (Often)

- Includes extensive information on clients, program activity, etc.
- Includes formats that support numerical processing (Excel, Access, etc.)
- Has extensive *metadata* describing the provenance of the information

Challenges (Often)

- Is collected for purposes other than evaluation
- Data do not relate to the evaluation indicators
- Data entry inconsistency

Administrative Files – Case Study

Summative Evaluation of the Labour Market Development Agreements (Part 2)

- Part 2 of the *EU Act* specifies eligibility and delivery of Employment Benefits and Support Measures (EBSMs)
- Goal of the evaluation was to measure the net impact of EBSMs on employability and employment post EI stint
- EI records used to identify recipients (under supervision of CRA) and extract data on Part 1 history (EI stints) and Part 2 participation
- Identify participants and non-participants in Part 2
- HRSDC supplied participant and non-participant sample contact information, as well as a scrambled SIN, to support client survey
- Client survey data (collected by PRA) returned to HRSDC which added EI history, EBSMs, and tax data and then returned the amalgamated file to PRA (scrambled SIN used to join data)
- PRA never received files with client ID or SIN

Collecting/Processing Quantitative Data (2)
Sample surveys

Key challenges:

- Sample frame definition and size (who and how many to contact)
- Survey mode (telephone, mail, online, hybrid)
- Question development
- Coding and analysis
- Align survey questions to evaluation questions

Create a questionnaire plan

Questions 101 – The basic structure

Program or Market Theory
Foundations of the program or marketing initiative

These are the main issues in the study as defined by the evaluation or marketing goals.

Measurement Theory
Operationalization of the abstract concepts

These are specific questions, often found in the evaluation framework.

Measurement Design
Specification of questions

These are the actual questions and response categories.

If all the problems of question wording could be traced to a single source, their common origin would probably prove to be in taking too much for granted.

S. Payne, *The Art of Asking Questions*, 1951

Question phrasing ‘rules’

- **Set wording to the respondent —**
  - Flesch-Kincaid Grade 8 - general population and
  - Grade 11 for civil servants and professionals
  - Technical jargon and acronyms with specialized audiences

- **Short sentences** in the active voice

Question phrasing 'rules' (cont'd)

- **Avoiding leading the respondent**
  - "*To balance the budget, should government reduce spending on the arts or raise taxes?"* is potentially a biased question.
  - Instead, use two questions:
    - *Do you agree or disagree that the budget should be balanced*?
    - [If agree] *Should government reduce spending on the art or raise taxes on socialists?*
  - *Better still use question reversal and a split ballet*

    a. *Do you agree or disagree that the budget should be balanced*

    Pose those who agree….

    a. *Some people support increasing taxes while others support reducing spending on arts in order to balance the budget. What is your preference?*

    b. *Some people support reducing spending on arts while others support increasing taxes in order to balance the budget. What is your preference?*

Framing questions and cognitive interviewing

Classic study (Belson, 1981)

- In a face-to-face interview, respondents were asked to agree or disagree with a series of statements such as "television shows are too violent for children."
- After carefully recording the responses, respondents were approached the day after and "debriefed" about the survey.
- Interviewers asked the respondents what they meant by "television shows," "too violent," and "children."
- These terms meant different things to different people, and he detected distinct meanings.
  - Television show meant primetime to some, and all times to others.
  - Children meant under 6, under 12, and under 18, depending on the respondent.
  - Too violent had meanings specific for each individual.
- With three meanings of child and two meanings of TV show, there are six questions being asked, let alone the infinite shades of "too violent."

Framing

- Many questionnaires make demands on memory.
- Framing practices
  - Use introductions and questions to set the stage.
  - Send a letter in advance explaining the survey and reminding the respondent of key dates
  - Avoid asking detailed questions about events or states in the past.

Four years ago, in 2003, how much money did you normally make at your job in a week, before taxes? If you did not have a job, please write "zero" or the number "0". Amount made per week $_______________

Poor

Please think about your current job (or the job held most recently). What is (was) the hourly wage $___________. When did you start that job ______. When did you leave that job (if currently employed) ________. Now please think about the job you held before that (Repeat sequence)

Better – CATI supports skips and branches to align questions to respondents

Two common biases

- ***Inter-item contamination***

  Qa. In your view, is AIDS a threat to someone who is heterosexual and not a drug user?

  Qb. Is the government providing sufficient funding to basic research in health?

Qa Contaminates Qb

- ***Social desirability bias***

  Qc. Have you heard of the XYZ program?

  Qd. In order to assess how well we are promoting our services, please tell me whether you have heard of the XYZ program.

Challenges

Shifts blame and allows someone to admit ignorance

Integrate the survey into the evaluation

- Survey questions must link to evaluation indicators
- The survey questionnaire aligns with the target sample (clients, stakeholders, managers, etc.)
- Create a “logical flow” (not the evaluation matrix flow!) to support respondent participation

Insert

- chtClient_survey_instrument_flow_chart_CM-LMAP.docx

Survey mode

- Telephone
- Mail
- Web/Online
- Hybrid – multimode

Telephone surveys

- Modern implementation uses computer-aided telephone interviewing (CATI)
- General population sampling uses variations of random digit dialing (RDD)
- Evaluation uses client-supplied contact information
- CATI typically embeds in a hybrid survey (mail-CATI; email-CATI)

Advantages

- Quick contact with respondents
- Interviewers can probe for more detailed/accurate answers
- Most effective method for surveying general population
- Works well as support to mail and online surveys (follow-up converts to interview)
- Supports skip logic

Disadvantages

- Response rates are usually lower
- Window of response is shorter
- Cannot support visual and other auditory messages
- Requires interviewer training and logistical management of call-backs
- Sample frames have limited contact information

Mail

- Supporting letter and questionnaire design “sell” the survey
- Name the respondent to ensure accurate delivery
- Personalization, layout, and design are critical features

Advantages

- Sample frames can be more easily verified
- Good design supports a “professional” look
- Visual cues support responses
- Allows respondents to set the questionnaire aside and return later
- Tends to work better with business and senior managers

Disadvantages

- Response rates are usually lower and require supplementary reminder (CATI)
- Skip logic is limited
- Survey is extended

Online

- Presumed to be "best" because it promises to be quick
- Appears to combine the benefits of telephone and mail modes
- Various providers offer a "turnkey" capacity to design, execute, and analyse (Survey Monkey)

Advantages

- Quick contact with respondents
- Supports skip logic
- Allows embedded designs, visuals, multi-media
- Client independence to design, execute, and analyze

Disadvantages

- Sample frames are often poor (email addresses outdated)
- Firewalls limit penetration, especially government and institutions
- Requires management of re-contacts
- Limited capacity to undertake non-response analysis

Hybrid

- Recommended for evaluation
- Letter/email contact with embedded link and PIN
- Telephone follow-up to non-respondents using CATI to
  - Remind
  - Collect data on the spot

Assessing Survey Quality

- **Item and overall non-response analysis**
- **Item non-response**
  - Identify questions where non-response is more frequent
  - Data replacement using
    - Averages
    - ICE (imputation by chained equations) and other regression methods
- **Total non-response**
  - Are non-responders different from responders?
  - Requires respondent attributes be attached to the sample frame
  - Use statistical weighting when discrepancy between sample and sample frame is too large

What is too large???

Turnkey online services limit callbacks and response rate enhancement, which is the central strategy to improving survey quality

Collecting/Processing Quantitative Data (1) Secondary analysis

- Most programs attempt to collect performance data
  - Client contact information
  - Billing data (health system)
  - Reports provided by project proponents (G&C)
- Performance (results) templates.

Creating a results information system database

- Follow the results chain to create a template
  - Inputs/activities (planned, actual, variance/reasons/mitigation)
  - Outputs (planned, actual, variance/reasons/mitigation)
  - Outcomes (immediate) (evidence/source/assessment)
  - Outcomes (intermediate) (evidence/source/assessment)
- Create data platform
  - Semi-annual reports in word
  - Access data reporting system

Insert

- Project Outcomes Summary_AB.xlsx
- Tab Results _ ALBERTA.docx

Collecting/Processing Qualitative Data (2) Interviews

- Sampling is **never** based on random selection
- Always choose respondents for their likelihood of providing relevant information
- Align respondents to evaluation issue
  - Senior → relevance/alternatives/management
  - Line manager → delivery/outcomes
  - Stakeholders → relevance/outcomes
- Interview guides should be focussed/concise
- **Discussion guide *not* questionnaire**

Insert

- ..\..\..\..\..\..\PROPOSAL\TEMPLATE\TECHNOTE\fgfindings_e.doc
- ..\..\..\..\..\..\PROPOSAL\TEMPLATE\TECHNOTE\indepth_e.pdf
- gdeGames_PCH Management_4_EN.doc
- Summative Evaluation of the Sport Funding Programs at Canadian Heritage.docx