Asking Questions

4 November 2009
Lecture 3

Topics for Today

- Designing Questions to Gather Factual Data
  - Question Objectives
- Definition of Concepts and Terms
- Knowing and Remembering
  - Do Respondents Know the Answer?
  - Stimulating Recall
  - Placing Events in Time
- The Form of the Answer
- Reducing the Effect of Social Desirability on Answers
  - Data Collection Procedures
- Source: Fowler Chapter 4

Questions to Gather Factual Data

- Designing questions which measure factual data
  - Quantifiable and verifiable (in theory)
  - There is some right answer (though the person might not know it)
- Some examples:
  - How many times they have used a given system in the past day/week/month
  - How many calls they have made to the help desk
  - The average response time for the IT staff to a given problem
  - The number of forms and buttons they need to fill out or press to accomplish a task
  - The number of hours of training they have received for a given technology

Goals for Factual Data Questions

- Sometimes we want to classify people:
  - Who is a heavy user of the system? Who is a light user?
  - What kind of workers find the existing system a burden to use
  - Which users use a specific application the most
- Or to count phenomena:
  - How many user interactions does the server see every day?
  - How many times does the application crash in the course of a week?
  - How many concurrent sessions are open with remote hosts from a given router at a time?
- Or both: By counting the number of application crashes per user per day, we can estimate
  - Reliability: the total number of crashes
  - Percentage of users who experience application crashes in a given day

Important because

It's important to classify and be aware of the goal of the question being asked because that can have a bearing on how the question should be asked and phrased

5 Challenges to Writing Good Questions

1. Defining objectives and specifying the kind of answers needed to meet the objectives of the question
2. Ensuring that all respondents have a shared, common understanding of the meaning of the question.
3. Ensuring that people are asked questions to which they know the answers. Barriers to knowing the answers can take at least three forms:
   a) never having the information needed to answer the question
   b) having the information at some point, but being unable to recall the information accurately or in the detail required by the question
   c) (for those questions that ask about events or experiences during some period of time) difficulty in accurately placing events in time
4. Asking the questions that respondents are able to answer in terms required by the question.
5. Asking questions the respondents are unwilling to answer accurately.

Example: Challenge 1

1. Defining objectives and specifying the kind of answers needed to meet the objectives of the question

- We may ask: "Does the application ever crash with unhandled exceptions? If so, when?"

- Fails the first test because:
  - Doesn't define the terms
  - How the respondent should answer
    - "Yes. Once or twice a week"
    - "Yes. On Sundays before 8am when I try to open up reports from the previous week's complete jobs and press on the 'Print to PDF' button."

Example: Challenge 2

2. Ensuring that all respondents have a shared, common understanding of the meaning of the question.

- We may ask: "Does the application ever crash with unhandled exceptions? If so, when?"

- Fails the second test because:
  - What if it's asked to non-technical people?
  - What's an "unhandled exception"?
  - Unhandled exceptions are shown differently in OS' and even versions of Windows
    - Hang, dialog box, error notice, debug notice

Example: Challenge 3

3. Ensuring that people are asked questions to which they know the answers. Barriers: never, unable to recall, can't place

- We could ask: "How many times in the past week were you forced to restart the application due to an error or it stopping to respond?"

- Here we fail all of them:
  - If you use it so often, who's counting?
  - They used to count, but now can't recall it
  - Can't remember what was last week, two weeks ago, or last month

Example: Challenge 4

4. Asking the questions that respondents are able to answer in terms required by the question.

- We could ask: "How many times in the past week were you forced to restart the application due to an error or it stopping to respond?" and let them fill in a number

- Here we fail the fourth principle:
  - What if they want to answer: "It crashes only when I have MS Word running in the background."
  - That's much more informative – but they can't write it

Example: Challenge 5

5. Asking questions the respondents are unwilling to answer accurately.

- We could ask: "What other applications were you running on your computer when the application last crashed?"

- Here we fail the fifth principle:
  - The answer might be a program the user is not supposed to be running at work (Solitaire, Kazaa, Skype, online games)
  - The interaction is important – but no one will admit it

Question Objectives

- The hardest part of writing questions is going from a simple question objective to a set of words which will accomplish the goal

- For example: Age.
  1. How old were you on your last birthday?
  2. On what date were you born?

- The first one is a bit unclear
  - Do you want exact age or a category (30s, 40s)?
  - May offend some people who are sensitive about their age.
    - Writing your birthday is a bit less (don't forget the year)
    - May open you up for discrimination issues

More Complex Objectives

- The more complex the issue, the more complex the question gets.
- For example: Income
  1. How much money do you make per month on your current job?
  2. How much money did you make in the last twelve months from paid jobs?
  3. What was the total income for you, and all family members living with you in your home, from jobs and from other sources of income during the last calendar year?
- What's the real objective? What do you really learn?
  - First one: job quality/rank
  - Second one: more accurate for a long term view of well-being or affluence
  - Third one: Bigger picture of income (but is a lot more invasive)

The Lesson

- The lesson is that when writing such questions it's essential that the researcher be aware precisely what kind of analysis will be done on the data gathered.

Example: Soft Drinks

- Soft Drink Consumption:
  1. How many soft drinks did you drink yesterday?
  2. How many soft drinks did you drink in the past week?
- What's the objective? What do we learn?
  - First one will be more accurate since people won't forget that easily, but is only a shorter period
  - Second one is over a longer period and:
    - More representative
    - Introduces more error since more people may make mistakes about how many they drank.

Example: Soft Drinks

- Soft Drink Consumption:
  1. How many soft drinks did you drink yesterday?
  2. How many soft drinks did you drink in the past week?
- Clarity: What's a soft drink?
  - Juice? Spring? Jump? Coca Cola?
  - Is a can the same as a 1.5L bottle? One cup?
- What's the goal of the analysis?
  - Get an accurate idea of consumption in general? (1st)
  - Measure soft drinks as a percentage of individual's diet? (2nd)

Soft Drinks and IS

- There are very clear parallels here for IS related subjects such as:
  - Usage of a particular application
  - Number manual steps you must perform which could be automated,
  - etc.

Example: Medical Care

- Use of Medical Care:
  1. How many times have you seen or talked to a doctor about your health in the past two weeks?
  2. How many times have you received any kind of medical care in the past two weeks?
  3. How many times have you received any kind of medical care in the last 12 months?
- Question is complicated since it's not clear what counts as medical care:
  - Does seeing a chiropractor count as medical care? How about a nurse?
  - What about doctors who are not your family doctor - eye doctor, psychiatrist?
- Period of time is important to
  - Classify people as heavy or light users of medical care
  - Count the number of incidents

Example: Medical Care

- Use of Medical Care:
  1. How many times have you seen or talked to a doctor about your health in the past two weeks?
  2. How many times have you received any kind of medical care in the past two weeks?
  3. How many times have you received any kind of medical care in the last 12 months?
- Third one filters better, but less accurate
- First and second are better to count incidents
- This has direct parallels for things such as
  - Number of help desk uses,
  - Number of times a help program is used

Question Objective List

One the best things that a researcher can do is to prepare a detailed list of research objectives before writing questions.

Example: Objective List

- Purpose of Survey: Study correlates of use of medical care. We think medical care is likely to be a function of the following:
  - Fiscal resources to afford medical care
  - Need for medical care
  - Access to medical care
  - Perception of value of medical care
- Within each of these categories, measurement objectives include:
- Fiscal resources relevant to medical care:
  - Annual family income past year (all sources)
  - Liquid assets (savings, bank accounts)
  - Health insurance
- Need for medical care
  - Chronic health conditions that might require care
  - Onset of acute illness
  - Injuries
  - Age/gender (to match with appropriate routine tests and exams
- Access to medical care
  - Regular provider or not
  - Perceived proximity of providers
  - Perceived ease of access
  - Perceived financial barriers
- Perception of value of medical care:
  - When not ill (checkups, screening, etc.)
  - For chronic conditions (not life-threatening)
  - For acute conditions (self-limiting)
- Use of medical care:
  - Visits to doctors
  - Other medical services (not M.D.)
  - Emergency room use
  - Hospitalization

Why Question Objective List?

1. The goal is to make sure you have fleshed out the topics and angles to ask
   - Has the benefit that as you write questions you can compare them against the objectives
2. Forcing you to think out what the survey's goals are, you can better think through what the goals should be and discover weaknesses
3. The exercise reminds you that designing the objectives is a separate task from the writing of the questions
   - The objectives list has no questions in it
   - The worst kind of surveys just take a simple objective and phrase it as a question
     - "Is the Application X valuable to your team?"
     - Assumes the respondent will come up with all of the structure for his answer and figure out goal of the question himself

5 Challenges to Writing Good Questions

1. Defining objectives and specifying the kind of answers needed to meet the objectives of the question
2. Ensuring that all respondents have a shared, common understanding of the meaning of the question.
3. Ensuring that people are asked questions to which they know the answers. Barriers to knowing the answers can take at least three forms:
   a) never having the information needed to answer the question
   b) having the information at some point, but being unable to recall the information accurately or in the detail required by the question
   c) (for those questions that ask about events or experiences during some period of time) difficulty in accurately placing events in time
4. Asking the questions that respondents are able to answer in terms required by the question.
5. Asking questions the respondents are unwilling to answer accurately.

Definition of Concepts and Terms

- Second issue is ensuring respondents all have a common understanding of the question
- Colloquialisms get in the way of clarity and should be defined
- For example: "How many computers do you have?"
  - Just desktops and laptops, not PDAs or smart phones?
  - All computing devices?
  - Playstation?
- If the point of the question is to count some categories more carefully:
  - Provide complete definitions so that all of most of the ambiguities about what is called for are resolved
  - Respondents can be asked to provide all of the information needed in order for the researcher to properly classify events for respondents.

Definition of Concepts and Terms

- For example: "How many computers do you have?"
- First way is more common – "Excluding/Including PDAs and Smartphones, how may computers do you have?"
  - Divides the question into different parts
  - Makes the question clearer
- Beware if we take it to an extreme.

Too Detailed

- Be very specific:
  - Next we need to get an estimate of the total income for you and family members living with you during 2008. When you calculate income, we would like you to include what you and other family members living with you made from jobs and also any income that you or other family members may have had from other sources, such as rents, welfare payments, social security, pensions, or even interest from stocks, bonds, or savings. So, including income from all sources, for you and for family members living with you, how much was your total family income in 2008?
- Much clearer, but involves a very complicated definition which may lose the answerer before he even gets to begin answering it
  - It also misses asking about pre-tax versus post-tax income and family composition

A Better Way: Catch Up

- To communicate complex definitions, it's better to try a different approach - ask follow up questions to filter responses
- For Income: Use a question series:
  - When you gave me the figure for your total family income, did you include any income you might have had from interest on stocks, bonds, or savings accounts?
  - When you gave me your income figure, did you include all the income that you had from rents?
  - Now, if you add in the kind of income you just mentioned that you did not include initially, what would be your estimate of your total family income in 2008?
- The catch-up question strategy can work well to catch commonly excluded items or answers

A Better Way: Break it Down

- Where definitions are very complex, make the questions just a series of questions about the parts desired and then assemble the sum total from all of the parts
- For Income, ask "What is your income?" and then:
  - What was your total family income from salaries alone in 2008?
  - What was your total family income from stocks, bonds, and savings accounts?
  - What was your total family income from rents?
- This way the researcher can assemble a total income picture as they want using the parts

Break it down vs. Catch Up

- Break-it-down has a few advantages:
  1. Makes the questions clearer since you don't need to communicate a complex definition
  2. Makes the reporting and recording much simpler
  3. May enable multiple measures combining different details
  4. Also makes the interview a bit longer, so the person has more of a chance to remember things

Impossible to Define

- Some definitions may be so complicated that the person may not even be aware of the distinctions.
- For example: "When you telecommute, what kind of remote login do you use - Virtual Private Networking, tunneling under SSH, Remote Desktop tunneled under SSL, Secure VNC?
- Many workers will not know of the answer
- Explaining the differences between them is beyond the scope of a simple interview or questionnaire.

Getting Around Impossible

- An alternative: ask specific questions about the topic to elicit details that can guide you towards the true answer or at least provide enough details to answer the question you are interested in
- For example:
  - When you log in remotely, do you see your desktop at your desk computer?
  - Do you need a special login token or password for remote login?
  - When you are remotely logged in, can you view everything on the work network or just one computer's desktop?
  - From these questions we can't necessarily answer the original one, but we can get a better idea of the situation, more so than by asking just a bunch of acronyms.
- Another approach: ask the user to narrate the answer to a complex question and then let the researcher go over the answer and extract the details that interest them
  - For example: Ask the users to narrate the steps they use to login remotely
  - Would hopefully still be reasonably long and could be decoded later

Challenge 2: Conclusions

- Essential that researchers and interviewees have a shared understanding of terms
- Doing so requires the researcher to think carefully about the definitions and what interests him
  - Some researchers try to include complex definitions in questions.
    - It can work for some definitions
  - More complex ones can be introduced using a general question followed by specifics
  - Another way is to try asking about details to do filtering
  - Finally, in case of complex definitions that the answerer might not know, the questionnaire can ask for limited details and try to extrapolate

5 Challenges to Writing Good Questions

1. Defining objectives and specifying the kind of answers needed to meet the objectives of the question
2. Ensuring that all respondents have a shared, common understanding of the meaning of the question.
3. Ensuring that people are asked questions to which they know the answers. Barriers to knowing the answers can take at least three forms:
   a) never having the information needed to answer the question
   b) having the information at some point, but being unable to recall the information accurately or in the detail required by the question
   c) (for those questions that ask about events or experiences during some period of time) difficulty in accurately placing events in time
4. Asking the questions that respondents are able to answer in terms required by the question.
5. Asking questions the respondents are unwilling to answer accurately.

Knowing and Remembering

- We have a question which everyone understands uniformly, now we are concerned whether the people are able to answer the question
- Three areas of difficulty:
  - The respondent may not have the information needed to answer the question
  - The respondent may once have known the information but have difficulty recalling it
  - For questions that require reporting events that occurred in a specific time period, respondents may recall that the events occurred but have difficulty accurately placing them in the time frame called for in the question
- We'll talk about these next

Do Respondents Know the Answer?

- Often questions are phrased to ask not just about a particular person's experiences, but to reflect information they know about others as well
- For example:
  - Asking one person about other members of his household
  - Asking a team leader about the experiences or needs of his team members
  - Asking one worker about his coworkers
- When such questions come up, can the person asked accurately report for others (reporting by proxy)?
  - Scientific studies that shows that reporting by proxy can be as accurate as self-reporting, but only for relatively public events and knowledge.
  - Asking about subjective states and preferences of others may be misleading or simply inaccurate

Unfamiliar Terms

- Another issue: asking people for technical terms which they are not familiar with
  - They may know the fundamental descriptions of what occurred, but are not aware of a technical term which may more specifically describe it
- For example: Users note that their application is running slowly or hangs a lot, but not know the meaning of the word "thrashing", "unhandled exception", or "virtual memory."
  - In some cases the user could describe the fundamental issue (*e.g.* "The program stopped and a little window popped up.") without necessarily knowing the technical term

Misconceptions

- Sometimes the respondent may have a misconception about the answer to the question
- **Example:** A new application is being designed and the number of expected users is being estimated.
- The person ordering the system may be aware of the number of target users but not be aware of their turnover rate or whether the target department is planning to significantly increase hiring soon.
- This is a very important issue when the number of terminals must be decided and the number of hours to be allocated for user training.
  - If there are 100 expected users but the turnover rate is 20% per year, there will be a need to allocate budget and time for training 20 new users per year. If the system was designed for 90-110 users and suddenly the number jumps to 200, the design criteria might have changed significantly

Do Respondents Know the Answer?

- There is no way around these issues, but sometimes having the wrong answer is worse than having no answer at all
  - At least it will be a "known unknown" instead of an "unknown unknown"
- How can we get people to recall information?

What do you recall?

- The brain is an amazing tool which doesn't completely forget anything
  - Certain things are easier to recall than others
- Three principles:
  1. The more recent the event, the more likely it is to be recalled
  2. The greater the impact or current salience of the event, the more likely it is to be recalled
  3. The more consistent an event was with the way the respondent thinks about things, the more likely it is to be recalled

Lesson: Asking about insignificant events, the window of memory is very short

Example: Asking about eating and soft drink consumption is accurate only for 24 hours
  - Asking for a longer period, people give an estimate based on their remembered "average" behavior rather than try to really recall
  - For a real measure, ask them to keep a diary

Example: Crime/Health

- **Example:** The US National Crime Survey and National Health Interview Survey would ask about crime and health events for the previous year.
- They noticed a huge drop off in accuracy after 6 months and so now only ask about a shorter period.
- The National Health Interview Survey only asks about the previous 2 weeks due to accuracy worries.

User Perspective

- From the user perspective, surveys are quick question-and-answer lists, so the respondent tries to answer it quickly
  - Most respondents will not invest a lot of energy in producing accurate responses
- How can we get users to recall better and think harder?
- One option: Make the question longer.
  - Doesn't mean making it complex - just add some introductory filler to make the question bigger
  - It forces the reader to think a bit longer about the question, increasing the chance of recall
  - It also may make user focus a bit more on the question

Other Recall Improvers

- Or: Ask multiple questions about the same topic
  - Allows the respondent to put more energy into recalling the event or item in question
  - Also may let the respondent think a bit more in detail and discover mistakes
- Or: Ask questions about events or items which are likely to be related to the desired question
  - May stimulate recall in the respondent.
- Or: Ask directly about commonly forgotten events
  - Allows respondents to recall them directly
- Or: Ask about the consequences of an event or item

Stimulation Example

- **Example:** "Did you ever need to restore your computer from backup?"
  - Will help recall crashes

Limits of Recall

- There are limits to recall, so if accuracy is required - develop a series of questions and correlated questions to get an accurate answer to a given question
- Aside from forgetting, there is sometimes to issue of over reporting, particularly for socially desirable behavior

False Recall

- Asking people the whether they voted always produces inaccurately high numbers of voters
  - Voting is considered a good thing to do, so people assume they always do it
  - Also asking people to recall *not doing something* is much more complex than the other around
- For false reporting, asking correlated questions and follow up to answers can help respondents filter out inaccurate responses.
  - For voting, reminding the respondents who was on the ballot
  - Also asking about related events - how you got to the voting place, where it was, who watched your kids

5 Challenges to Writing Good Questions

1. Defining objectives and specifying the kind of answers needed to meet the objectives of the question
2. Ensuring that all respondents have a shared, common understanding of the meaning of the question.
3. Ensuring that people are asked questions to which they know the answers. Barriers to knowing the answers can take at least three forms:
   a) never having the information needed to answer the question
   b) having the information at some point, but being unable to recall the information accurately or in the detail required by the question
   c) (for those questions that ask about events or experiences during some period of time) difficulty in accurately placing events in time
4. Asking the questions that respondents are able to answer in terms required by the question.
5. Asking questions the respondents are unwilling to answer accurately.

Placing Events in Time

- Some surveys ask questions about events that have taken place in the past, often during a long period of time
  - Ex: How many times a person was hospitalized in the past twelve months
- The problem: People have a hard time remembering when events took place within an arbitrarily defined period of time
  - 1 week, 1 month, or 1 year
  - Even if they remember the events, they will have a hard time placing them in terms of a particular period of time
- Two approaches:
  - Stimulate recall activities on the part of the respondents to help them place events in time
  - Design data collection procedures that generate boundaries for reporting periods

Stimulating Recall

- You can try showing the respondent a calendar to help them think about when events took place
- You can have the respondent fill in or note important life events such as birthdays, weddings, or holidays can help them recall
- Ask them to think about their life one year ago: where they lived, where they were working, what was going on in the family
- If they have a vision of their life events, they can then associate it with a date on the calendar and will then have an easier time relating the question to dates
  - "Did you go to the hospital for your broken leg before or after Timmy's birthday?"

Stimulating Recall

- Another idea: Have the respondents recall associated events around the question
  - What was the weather like when you went to the hospital?
  - What were you wearing when you went to the hospital?
- This will take more time than a standard survey or interview might take and end up being very personalized for the respondent

Creating Boundaries

- Another idea: Create an artificial boundary by scheduling two or more interviews with the same person
- At the first interview, the respondent is told that at the next interview he will be asked about the events that took place since the current one
- Three advantages:
  1. Although the interview is not necessarily a big event in the person's life, it gives a clearly defined anchor in time to relate to
  2. If there are worries of double counting, the researcher can ask about events during the first interview and correlate them with the responses from the second interview. Cases where events are "telescoped" from before the first interview to the period between the two interviews can then be easily noticed
  3. The respondents being told that they will be asked about some events in the future makes them more aware of the events when they occur and more likely to recall them later
- This approach is more expensive in terms of time and money for all involved, but if it's essential to know information in a given time period, it can help

Diaries

- One last technique: Keeping a diary of events so that respondents can write down events
- Getting respondents to keep a diary is a problem by itself, but they can be useful if the respondent agrees
- They are especially useful in tracking diet and small expenditures for a short period of time

Summary

- Designing Questions to Gather Factual Data
  - Question Objectives
- Definition of Concepts and Terms
- Knowing and Remembering
  - Do Respondents Know the Answer?
  - Stimulating Recall
  - Placing Events in Time
- The Form of the Answer
- Reducing the Effect of Social Desirability on Answers
  - Data Collection Procedures