



ECE1724H S2: Empirical Software Engineering

Survey Design

Shurui Zhou

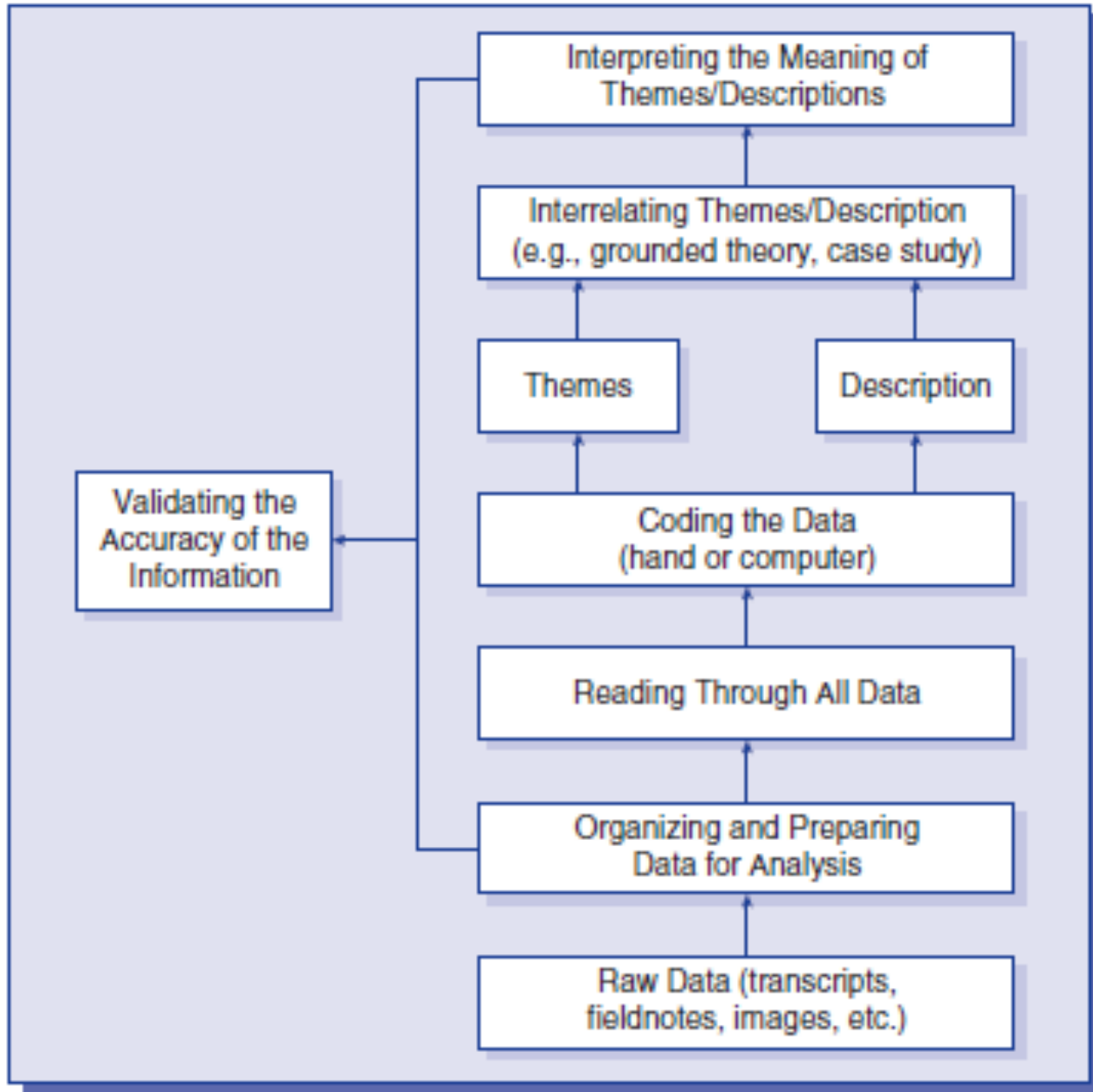


The Edward S. Rogers Sr. Department
of Electrical & Computer Engineering
UNIVERSITY OF TORONTO

Interview



Figure 9.1 Data Analysis In Qualitative Research



Data Analysis in Qualitative Research

Creswell. Research Design.

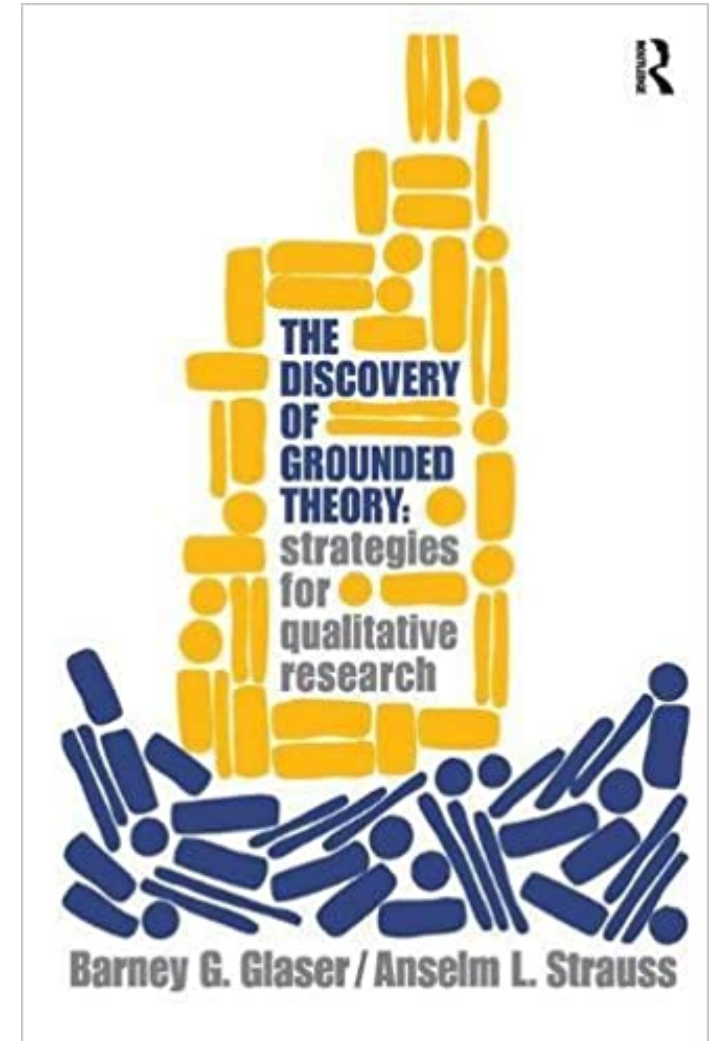
Grounded Theory

- Focus on generating theoretical ideas (or hypotheses) from the data
- rather than having these specified beforehand

“A grounded theory is one that is inductively derived from the study of the phenomena it represents.”

-- Strauss and Corbin. p23

**Key focus = reflective reading of text
and the application of codes**



Types of Coding - Examples

¹I asked the principal what the need for the new program was, and he responded that the students coming into the 9th grade were two years below grade level and that the old curriculum was ineffective. Through testing (the Nelson Reading Test) it was determined that students were growing academically only 5 or 6 months during the 10-month school year.

¹MOTIVATION

Challenges:

- To make sense of massive amounts of data
- To reduce the volume of information
- To identify significant patterns
- To construct a framework to communicate what the data reveals

Coding stages

- First Cycle
 - Three elemental method: Descriptive, In Vivo, Process coding
- Second Cycle

EXAMPLE

REACTION

It was really good. There was a variety of activities, the overhead and information where they talked about it, and the opportunity to practice the activities together. I liked it. The 5 hours went really quickly. We had a good group, and felt very comfortable because everyone was open and sharing. And the lunch was wonderful. Having lunch was a good idea.

STRUCTURE
ACTIVITIES

COMFORT LEVEL

Yes, the structure helped my grasp the information, and I enjoyed the group size and variety of activities. 5.5 hours was good enough, and it went quickly. It all seemed to follow their outline, and it gave the opportunity to listen and then practice and get to know other people, because I was there by myself.

STRUCTURE
VARIETY

SOCIAL NETWORKING

ONSITE
SOCIAL
ASPECT

I think I would have gotten the same information either way with the overheads and printouts, but the interpersonal and opportunity to relate to other people and have back and forth and the ability to ask questions was more personal and enjoyable. In a webinar, I don't feel comfortable asking questions to someone I don't know, so the personal, face to face was better than if I did the webinar. I imagine I would have gotten the same information, but it wouldn't have been as enjoyable, and without the activities and other people's questions, I probably wouldn't remember as much.

FACE-TO-FACE
VS

ONSITE IMPACT

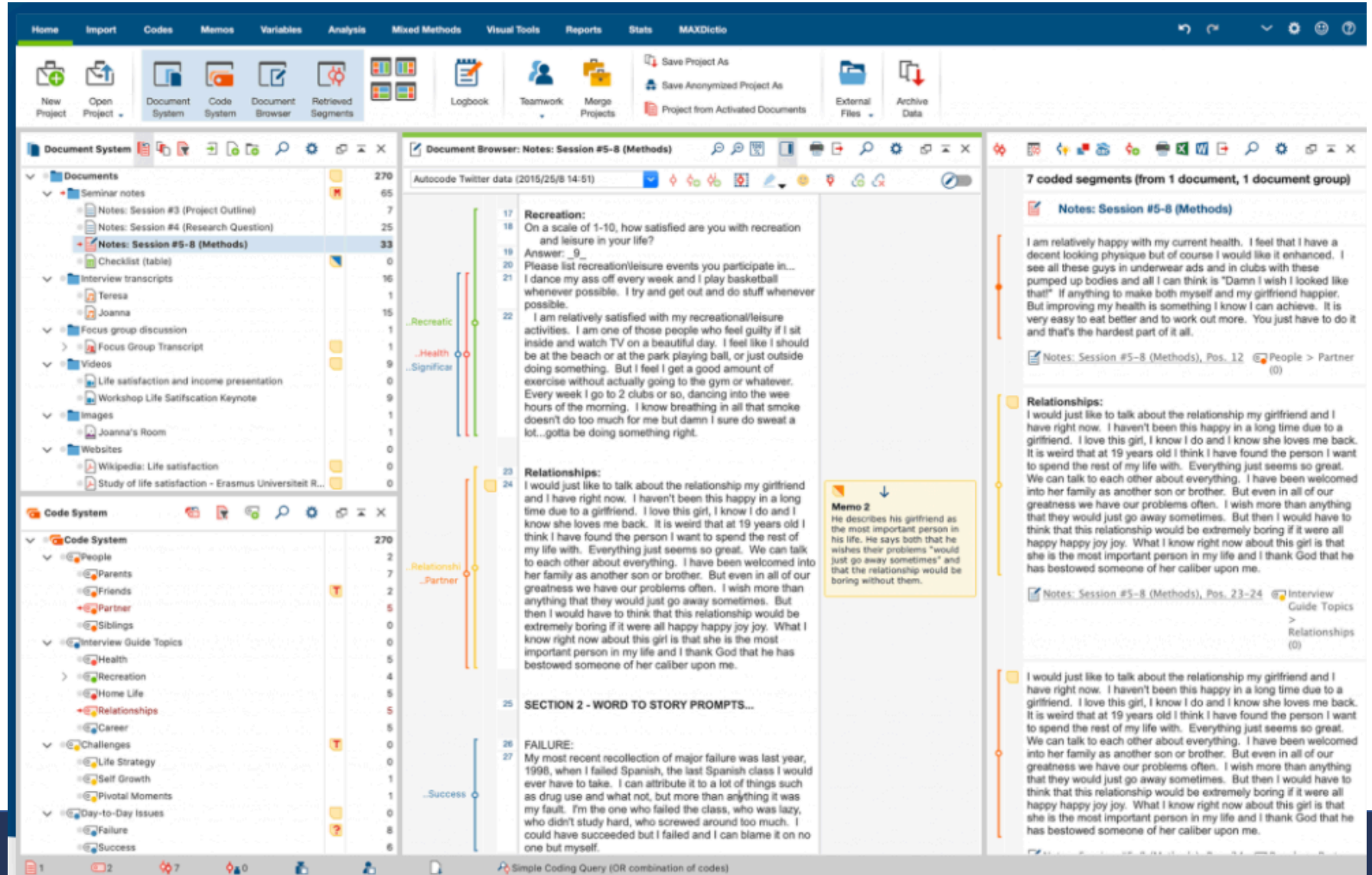
REMOTE

1. What was your overall impression of the training in terms of its usefulness to your work?

REACTION

It is great because I do survey development and work with people that develop surveys. It was really helpful. Afterward, I analyzed surveys and it made me wish I had attended the training before, because now I know there's better ways to do it. Yes, it has definitely met my expectations.

EXPECTATIONS



The screenshot displays the MAXQDA software interface, which is organized into several main panels:

- Top Menu Bar:** Includes Home, Import, Codes, Memos, Variables, Analysis, Mixed Methods, Visual Tools, Reports, Stats, and MAXIDictio.
- Toolbar:** Contains icons for New Project, Open Project, Document System, Code System, Document Browser, Retrieved Segments, Logbook, Teamwork, Merge Projects, Save Project As, Save Anonymized Project As, Project from Activated Documents, External Files, and Archive Data.
- Document System Panel (Left):** A tree view showing the project's structure. It includes folders for Documents, Interview transcripts, Focus group discussion, Videos, Images, Websites, and Code System. The 'Notes: Session #5-8 (Methods)' document is selected.
- Code System Panel (Bottom Left):** A tree view showing the hierarchical structure of codes. It includes categories like People, Interview Guide Topics, Challenges, and Day-to-Day Issues. The 'Relationships' code is selected.
- Document Browser Panel (Center):** Displays the content of the selected document, 'Notes: Session #5-8 (Methods)'. It shows a list of sections with corresponding line numbers and a vertical timeline on the left. The sections include:
 - Recreation:** On a scale of 1-10, how satisfied are you with recreation and leisure in your life? Answer: 9. Please list recreation/leisure events you participate in... I dance my ass off every week and I play basketball whenever possible. I try and get out and do stuff whenever possible. I am relatively satisfied with my recreational/leisure activities. I am one of those people who feel guilty if I sit inside and watch TV on a beautiful day. I feel like I should be at the beach or at the park playing ball, or just outside doing something. But I feel I get a good amount of exercise without actually going to the gym or whatever. Every week I go to 2 clubs or so, dancing into the wee hours of the morning. I know breathing in all that smoke doesn't do too much for me but damn I sure do sweat a lot...gotta be doing something right.
 - Relationships:** I would just like to talk about the relationship my girlfriend and I have right now. I haven't been this happy in a long time due to a girlfriend. I love this girl, I know I do and I know she loves me back. It is weird that at 19 years old I think I have found the person I want to spend the rest of my life with. Everything just seems so great. We can talk to each other about everything. I have been welcomed into her family as another son or brother. But even in all of our greatness we have our problems often. I wish more than anything that they would just go away sometimes. But then I would have to think that this relationship would be extremely boring if it were all happy happy joy joy. What I know right now about this girl is that she is the most important person in my life and I thank God that he has bestowed someone of her caliber upon me.
 - SECTION 2 - WORD TO STORY PROMPTS...**
 - FAILURE:** My most recent recollection of major failure was last year, 1998, when I failed Spanish, the last Spanish class I would ever have to take. I can attribute it to a lot of things such as drug use and what not, but more than anything it was my fault. I'm the one who failed the class, who was lazy, who didn't study hard, who screwed around too much. I could have succeeded but I failed and I can blame it on no one but myself.
- Memo Panel (Right):** Displays the content of the selected memo, 'Notes: Session #5-8 (Methods)'. It shows a list of sections with corresponding line numbers and a vertical timeline on the left. The sections include:
 - Recreation:** On a scale of 1-10, how satisfied are you with recreation and leisure in your life? Answer: 9. Please list recreation/leisure events you participate in... I dance my ass off every week and I play basketball whenever possible. I try and get out and do stuff whenever possible. I am relatively satisfied with my recreational/leisure activities. I am one of those people who feel guilty if I sit inside and watch TV on a beautiful day. I feel like I should be at the beach or at the park playing ball, or just outside doing something. But I feel I get a good amount of exercise without actually going to the gym or whatever. Every week I go to 2 clubs or so, dancing into the wee hours of the morning. I know breathing in all that smoke doesn't do too much for me but damn I sure do sweat a lot...gotta be doing something right.
 - Relationships:** I would just like to talk about the relationship my girlfriend and I have right now. I haven't been this happy in a long time due to a girlfriend. I love this girl, I know I do and I know she loves me back. It is weird that at 19 years old I think I have found the person I want to spend the rest of my life with. Everything just seems so great. We can talk to each other about everything. I have been welcomed into her family as another son or brother. But even in all of our greatness we have our problems often. I wish more than anything that they would just go away sometimes. But then I would have to think that this relationship would be extremely boring if it were all happy happy joy joy. What I know right now about this girl is that she is the most important person in my life and I thank God that he has bestowed someone of her caliber upon me.
 - FAILURE:** My most recent recollection of major failure was last year, 1998, when I failed Spanish, the last Spanish class I would ever have to take. I can attribute it to a lot of things such as drug use and what not, but more than anything it was my fault. I'm the one who failed the class, who was lazy, who didn't study hard, who screwed around too much. I could have succeeded but I failed and I can blame it on no one but myself.



ATLAS.ti - Qualitative Data Analysis

 ATLAS.ti CLOUD

Analyze Your Qualitative Data Online

ATLAS.ti Cloud: Analyze your qualitative data online

<https://www.youtube.com/watch?v=yjBN88HMbTs>

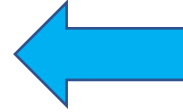




Question:
Why don't we use ML models
to code the transcript?

Agenda for Today


- Your Grounded Theory experiences
- Paper reading presentation
- Designing a survey





- - Razavi, M. N., & Iverson, L. (2006). A grounded theory of information sharing behavior in a personal learning space Preview the document, Proceedings of the ACM Conference on Computer Supported Cooperative Work (pp. 459-468).
- - de Souza, C. R., & Redmiles, D. F. (2008). An empirical study of software developers' management of dependencies and changes Preview the document, Proceedings of the 30th International Conference on Software Engineering (pp. 241-250).
- - Deterding, S. (2016). Contextual autonomy support in video game play: a grounded theory. Preview the document Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016.

Agenda for Today

- Your Grounded Theory experiences
- Paper reading presentation
- Designing a survey 



Survey Research

“A comprehensive system for collecting information to describe, compare or explain knowledge, attitudes and behavior over large populations”

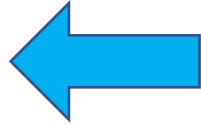


EXAMPLE

- Course evaluation
- Public opinion pollsters
- Current Population Survey to help set economic policy
- Market researchers -- consumer attitudes and behaviors
- ...

Main activities

- Setting the objectives
- Survey design
- Developing the survey instrument (i.e. the questionnaire)
- Evaluating the survey instrument
- Obtaining valid data
- Analysing the data



Shari Lawrence Pfleeger and Barbara A. Kitchenham, "Principles of Survey Research," Software Engineering Notes, (6 parts) Nov 2001 - Mar 2003

What is a Survey Good For?

What is a Survey Good For?

- Prevalence (sort of)
- Relations between variables
- Differences among subpopulations

Limitation?

Limitations

- Relies on self-reported observations
- Difficulties of sampling and self-selection
- Information collected tends to subjective opinion

When to use Survey Research?

- To evaluate the **frequency** of some characteristic across a population
 - E.g. how many companies use agile methods?
- To evaluate the **severity** of some condition that occurs in a population
 - E.g. what's the average cost overrun of software projects?
- To identify factors that **influence** a characteristic or condition
 - E.g. What factors cause companies to adopt new ASE tools?

Is survey an appropriate research method to address the stated objectives? (checklist)

- Is it clear what population can answer the survey questions reliably?
- Is there a method of obtaining a representative sample of that population?
- Do you have resources to obtain a large enough sample?
- Is it clear what variables need to be measured?
- Is it clear how to measure the variables?

What type of question are you asking?

Exploratory

→Existence:

- ↳ Does X exist?

→Description & Classification

- ↳ What is X like?
- ↳ What are its properties?
- ↳ How can it be categorized?
- ↳ How can we measure it?
- ↳ What are its components?

→Descriptive-Comparative

- ↳ How does X differ from Y?

→Frequency and Distribution

- ↳ How often does X occur?
- ↳ What is an average amount of X?

→Descriptive-Process

- ↳ How does X normally work?
- ↳ By what process does X happen?
- ↳ What are the steps as X evolves?

Correlation

→Relationship

- ↳ Are X and Y related?
- ↳ Do occurrences of X correlate with occurrences of Y?

→Causality

- ↳ Does X cause Y?
- ↳ Does X prevent Y?
- ↳ What causes X?
- ↳ What effect does X have on Y?

→Causality-Comparative

- ↳ Does X cause more Y than does Z?
- ↳ Is X better at preventing Y than is Z?
- ↳ Does X cause more Y than does Z under one condition but not others?

Causal Relationship

→Design

- ↳ What is an effective way to achieve X?
- ↳ How can we improve X?


Design

Base-rate

Just a
reminder...



Main activities

- Setting the objectives
- Survey design 
- Developing the survey instrument (i.e. the questionnaire)
- Evaluating the survey instrument
- Obtaining valid data
- Analysing the data

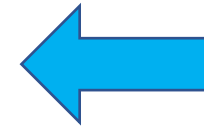
Shari Lawrence Pfleeger and Barbara A. Kitchenham, "Principles of Survey Research," Software Engineering Notes, (6 parts) Nov 2001 - Mar 2003

Survey Design

- **Cross sectional:** Used to obtain a snapshot of participants' current activities.
- **Longitudinal:** Administer a survey periodically to track changes over time

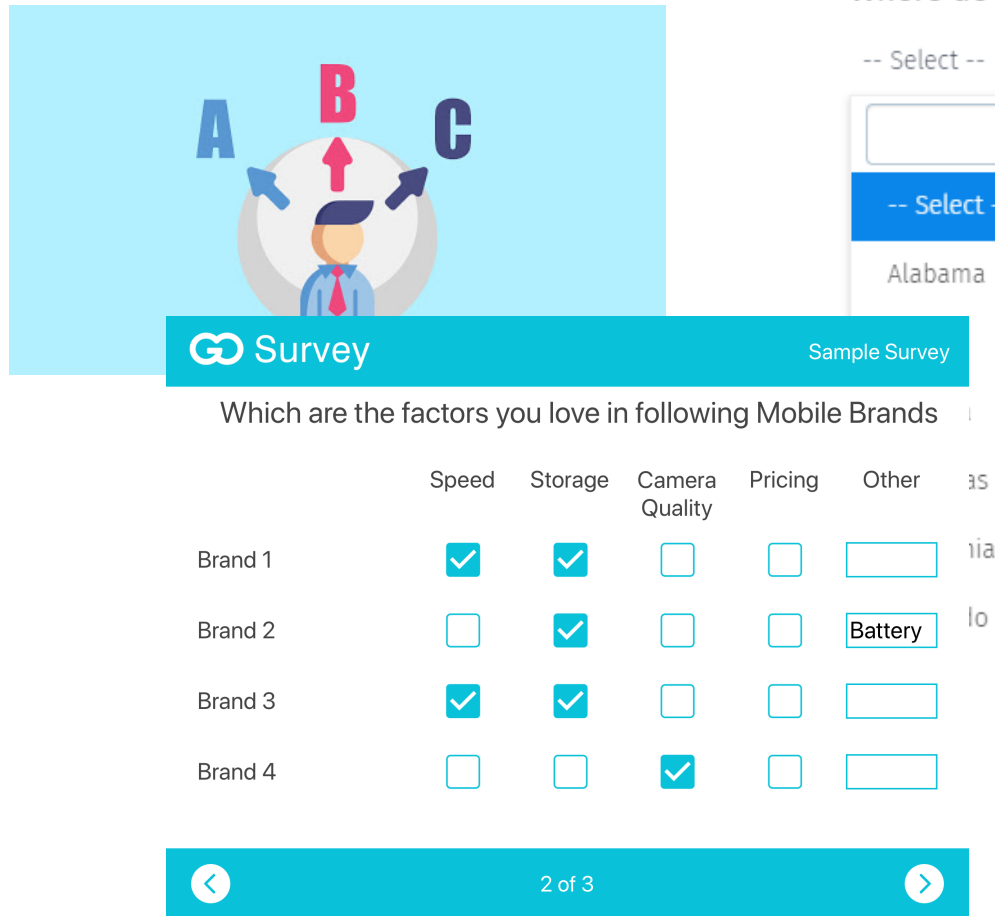
Main activities

- Setting the objectives
- Survey design
- Developing the survey instrument (i.e. the questionnaire)
- Evaluating the survey instrument
- Obtaining valid data
- Analysing the data



Shari Lawrence Pfleeger and Barbara A. Kitchenham, "Principles of Survey Research," Software Engineering Notes, (6 parts) Nov 2001 - Mar 2003

Closed-ended Questions



Survey Sample Survey

Which are the factors you love in following Mobile Brands

	Speed	Storage	Camera Quality	Pricing	Other
Brand 1	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Brand 2	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Battery
Brand 3	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Brand 4	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

2 of 3

Where do you live?

-- Select --

-- Select --

Alabama

Please rank the following features in order of importance, where 1 is the most important to you.

Location	1
Comfort	2
Service	3
Value for money	4

← COMPLETE

Closed-ended Questions

How likely are you to share your location to meet friends after work?

- Absolutely never
- Sometimes
- Occasionally
- Once or more a week
- Everyday



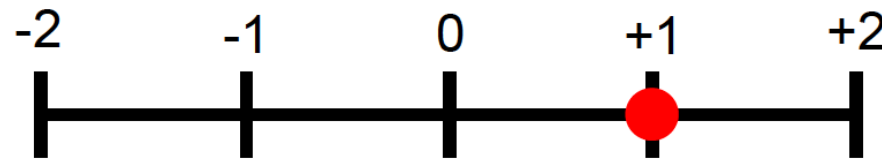
Is it easy or difficult to distinguish
between these three categories?

If difficult, why?

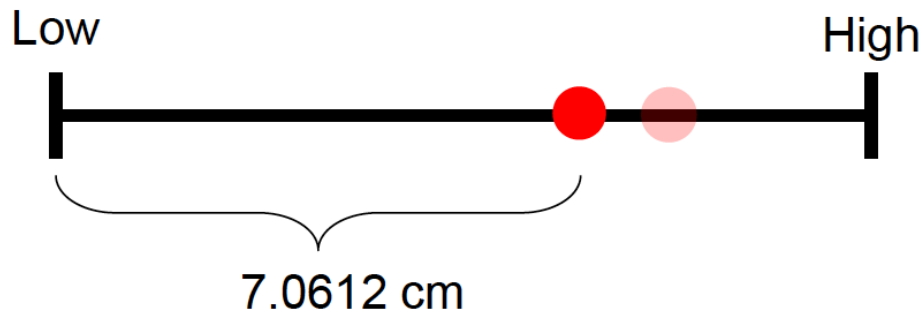
Ordinal Scales

- Ordinal or interval scales ask interviewees to choose a “level” of the variable of interest

Numbered Scale:
(choose your number)



Visual Analogue
Scale:
(mark your level)



How do you use an ordinal scale?

- *Understand nuanced opinions.* Do respondents “agree” or “strongly agree” with a stance on an issue?
- *Uncover perceptions.* Do respondents find a particular statement “false,” “mostly false,” “mostly true,” or “true”?
- *Measure relative performance.* Is a certain employee “more productive,” “just as productive,” or “less productive” than other employees?
- *Gauge sentiment.* Is a customer “very satisfied,” “satisfied,” “dissatisfied,” or “very dissatisfied” with a recent purchase?

EXAMPLE

Employee Satisfaction Survey Template

How meaningful is your work?

- Extremely meaningful
- Very meaningful
- Moderately meaningful
- Slightly meaningful
- Not at all meaningful

How challenging is your job?

- Extremely challenging
- Very challenging
- Moderately challenging
- Slightly challenging
- Not at all challenging

Open-ended Questions

Exploratory in nature, and offer the researchers rich, qualitative data. In essence, they provide the researcher with an opportunity to gain insight on all the opinions on a topic they are not familiar with.

Open-ended Questions

- **Definition and designation questions**

What-is asks to develop definitions of things

Who identifies the responsible agent

What-kinds-of ask for possible types and exemplars

- **Process, event and exception questions**

How-to ask how an action is performed

When asks about timing constraints, pre-and post-conditions

What-if asks about failures or unexpected events

Follow-on questions result from answers from previous questions

Open Ended Questions – Acquiescence

FIGURE 5.4 State both positive and negative sides in the question stem.

<u>Poor Designs</u>	<u>Improved Designs</u>
Do you favor congressional term limits of four years? <input type="radio"/> Favor <input type="radio"/> Oppose	Do you favor or oppose congressional term limits of four years? <input type="radio"/> Favor <input type="radio"/> Oppose
How satisfied are you with the overall service you have received from your financial consultant? <input type="radio"/> Very satisfied <input type="radio"/> Somewhat satisfied <input type="radio"/> Somewhat dissatisfied <input type="radio"/> Very dissatisfied	How satisfied or dissatisfied are you with the overall service you have received from your financial consultant? <input type="radio"/> Very satisfied <input type="radio"/> Somewhat satisfied <input type="radio"/> Somewhat dissatisfied <input type="radio"/> Very dissatisfied

EXAMPLE

- What do you like most about our new product?
- What changes would most improve our product?

1. How often do you attend events in this neighborhood?

- Extremely often
- Quite often
- Moderately often
- Slightly often
- Not at all often

2. If you do not attend events in this neighborhood, why not?



<https://www.surveymonkey.com/>



Products ▾

Solutions ▾

Resources ▾

Plans & Pricing

Log in

Sign up

Will my product be a success or a flop?

A global leader in survey software. 20 million questions answered daily.

Get started

Response Rate and Motivation

Which one is more motivating?

“The questionnaire is designed to discover what aspects of your educational background have been useful to you in your career. The results of the survey will be used to help improve curricula. All the information you provide will be kept confidential. In particular we have no intention of judging you as a person—we are merely interested in learning about the relevance of certain topics to your work.”

“Dear Executive, We are sponsoring a study for the University of X, and Professors Y and Z. It is only through our cooperative efforts with the academic community that we bring our commercial experiences to the classroom. Thank you for your help.”

How to Get People to Respond

- Make it easy
 - Keep it short!
 - Keep it simple and clear and convenient
 - Choice of modes may reduce response rates
- Make it seem valuable, important
 - How will results be useful?
 - Ask for help/advice
 - Make it interesting
 - Gamification
 - Scarcity – only some people get asked
 - Others have responded
 - Pay (a little bit) forward

How to Get People to Respond

- Be trustworthy
 - Worry about malware, fake surveys
 - Ways to assess your trustworthiness (contact info)
 - Sponsorship
 - Assure confidentiality
 - Minimize requests to obtain personal information
- Be likeable
 - Show respect
 - Don't use subordinate language

Length of a Survey?

How much time are respondents willing to spend on your survey?



Brent Chudoba
3 min read



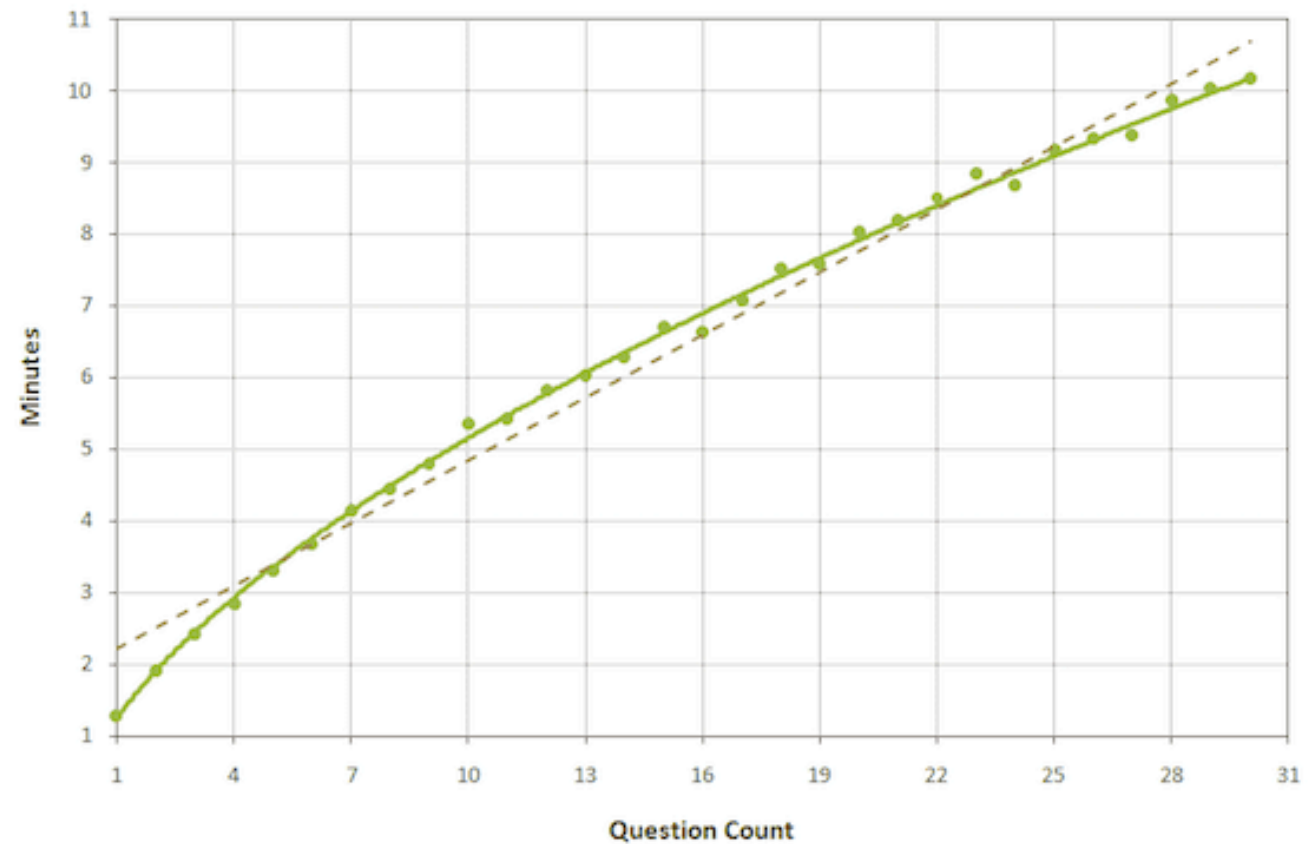
https://www.surveymonkey.com/curiosity/survey_completion_times/

Question Count	Average Seconds Spent Per Question*	Total Survey Completion Times
----------------	-------------------------------------	-------------------------------

1	75	1 min 15 sec
2	40	2 min
3-10	30	2 - 5 min
11-15	25	5 - 7 min
16-25	21	7 - 9 min
26-30	19	9 - 10 min

* Rounded and grouped for illustrative purposes

Response Times





Identifying Redundant PRs on GitHub



DuplicatePR-bot commented on Oct 17, 2019



Hi there! This pull request looks like it might be a duplicate of [#1370](#), since it has *the same issue number* , *a similar title*, and similar commits.

To improve our bot, you can help us out by clicking one of the options below:

- This pull request is a **duplicate**, so this comment was **useful**. [check](#)
- This pull request is **not a duplicate**, but this comment was **useful** nevertheless. [check](#)
- This pull request is **not a duplicate**, so this comment was **not useful**. [check](#)
- I do not need this service, so this comment was **not useful**. [check](#)

This bot is currently in its alpha stage, and we are only sending at most one comment per repository. If you are interested in using our bot in the future, please [subscribe](#). If you would like to learn more, see our [web page](#).



sergeyrolich commented on Oct 17, 2019

Contributor

Author



Duplicate [#1370](#), close



1

Consent Agreement

We're conducting this survey as part of a research project to find out problems and practices in fork-based development and how tool support can help to solve those problems. It should take no more than 5 minutes. Any participant should be at least 18. Your participation in this study is voluntary. We will just save your feedback to supplement our notes, but we will not release any information without anonymization. The records may be reviewed by the study sponsor, in this case NSF.

Thank you for your input! Here are two short questions:

Your response of #PR_1[link] might be duplicate with #PR_2[link]

- ☒ This pull request is a duplicate, so this comment was useful
- ☐ This pull request is not a duplicate, but this comment was useful nevertheless.
- ☐ This pull request is not a duplicate, so this comment was not useful.
- ☐ I do not need this service, so this comment was not useful.

If you'd like to help us out **even more**, please let us know below why this comment was useful to you...

Submit

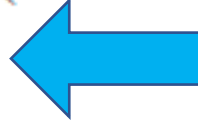
...and how we could improve:

Submit

Who are we?

We are researchers at [Carnegie Mellon University](#) developing a bot to monitor GitHub projects and alert project maintainers to potentially duplicate pull requests. [Shurui Zhou](#) is a fifth-year PhD student studying under Dr. [Christian Kaestner](#). [Annika Esau](#) is an undergraduate [REU](#) student.

Main activities

- Setting the objectives
- Survey design
- Developing the survey instrument (i.e. the questionnaire)
- Evaluating the survey instrument 
- Obtaining valid data
- Analysing the data

Shari Lawrence Pfleeger and Barbara A. Kitchenham, "Principles of Survey Research," Software Engineering Notes, (6 parts) Nov 2001 - Mar 2003

Pre-testing

- To check that the questions are understandable.
 - To assess the likely response rate and the effectiveness of the follow-up procedures.
 - To evaluate the reliability and validity of the instrument.
 - To ensure that our data analysis techniques match our expected responses.
- Focus groups
 - Pilot studies

Reliability

○ Test-Retest Reliability

- If the same person answers the survey twice, do you get the same answers?
- Problems:
 - What if the world has changed?
 - What if answering the questionnaire changes their attitude?
 - What if they just remember their answers from last time?

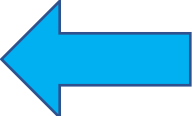
○ Alternate Form Reliability

- Do re-worded or re-ordered questions yield the same results?

○ Inter-rater Reliability

- If someone else administers the questions, do you get the same answers?
- If someone else codes the responses, do you get the same results?

Main activities

- Setting the objectives
- Survey design
- Developing the survey instrument (i.e. the questionnaire)
- Evaluating the survey instrument
- Obtaining valid data 
- Analysing the data

Shari Lawrence Pfleeger and Barbara A. Kitchenham, "Principles of Survey Research," Software Engineering Notes, (6 parts) Nov 2001 - Mar 2003

Sampling

I. Probability Sampling

- Simple random sampling
- Stratified random sampling
- Systematic random sampling

II. Non-probability Sampling

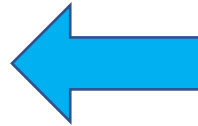
- Convenience sampling
- Purposive sampling
- Expert sampling
- Quota sampling
- Modal sampling
- Heterogeneity sampling
- Snowball sampling

Avoiding Sampling Bias

- Clear definition of the survey sample:
 - Define the U, the unit of analysis
 - Define the P, the target population
 - ...such that $P = \{U\}$
 - Sample of the entire target population
 - not just the most accessible portion of it!
- Stratified Random Sampling for confounding variables:
 - E.g. U = individual developer,
P = developers working in Canadian software companies
 - ... but what if 80% of our sample comes from a single, dominant company?
 - If we really wanted U = Canadian Software Companies
 - Then change P
 - Alternatively, if company is a confounding variable
 - Group population by company, then sample within each

Main activities

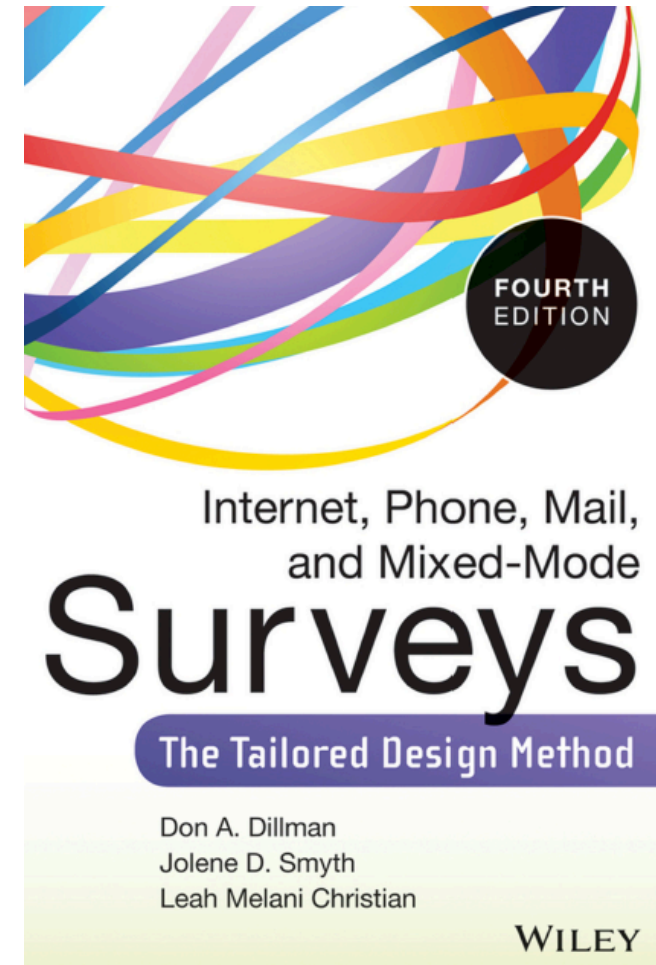
- Setting the objectives
- Survey design
- Developing the survey instrument (i.e. the questionnaire)
- Evaluating the survey instrument
- Obtaining valid data
- Analysing the data



Shari Lawrence Pfleeger and Barbara A. Kitchenham, "Principles of Survey Research," Software Engineering Notes, (6 parts) Nov 2001 - Mar 2003

References

- Chapters from Dillman, D., Smyth, J. D., & Christian, L. M. (2014). Internet, Phone, Mail and Mixed-Mode Surveys: The Tailored Design Method (4th ed.). Hoboken, NJ: Wiley.
 - Ch1: Sample Surveys in our Electronic World
 - Ch2: Reducing People's Reluctance to Respond to Surveys
 - Ch4: The Fundamentals of Writing Questions
 - Ch5: How to Write Open and Closed Ended Questions



EXAMPLE

- Ecosystem survey:
http://cmu.ca1.qualtrics.com/jfe/form/SV_d4M66VwPlZYd5kh
Results: <http://breakingapis.org/survey/>
- GitHub open source survey:
<https://github.com/github/open-source-survey>
Results: <http://opensourcesurvey.org/2017/>

