Empirical Software Engineering Research

Qualitative Studies

Norman Peitek, Annabelle Bergum, Lina Lampel, Sven Apel

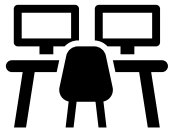
Learning Goals

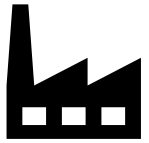
- Understand use cases for qualitative methods
- Learn opportunities and risks of interviews, case studies, and questionnaires
- Gain an overview over a typical qualitative data-analysis process

Qualitative Studies

- A qualitative study is conducting empirical investigations yielding qualitative data and where we apply qualitative data-analysis methods
- Remember, quantitative methods use numbers, while qualitative methods focus on words and concepts
- They are very prevalent in the social sciences as they focus on humans
- In SE, the human aspect is also important, so qualitative studies are useful in the human and social aspects of SE

Qualitative Studies





- Qualitative studies are typically based on
  - Interviews of programmers (or other stakeholders)
  - Observations of people at work (meetings, programming activities, …)
  - Analysis of archival data
- The overall process is similar (objectives, design, collect data, analysis, conclusion)
  - The difference is the degree of objectivity in the data analysis
  - Qualitative studies are more subjective due to the interpretation (and harder to replicate)
  - Provide evidence that the interpretation is sound and reasonable based on the data
- We study aspects of SE “in the field”

Quantitative versus Qualitative Studies

Quantitative Studies	Qualitative Studies
Numbers	Words (images, concepts)
Researcher-driven	Participant-driven
Researcher is distant	Researcher is close
Theory is tested against data	Theory emerges from data
Linear	Iterative
Structured	Unstructured
Generalization-oriented	Context-oriented
Hard, reliable data	Rich, deep data
Behavior	Meaning
Artificial setting	Natural settings in the field

Some quantitative may have no theoretical basis. Some qualitative studies may already have some pre-existing theory.

Reality is not as binary! Quantitative studies also can find some meaning in the data. Qualitative studies also look for patterns in behavior.

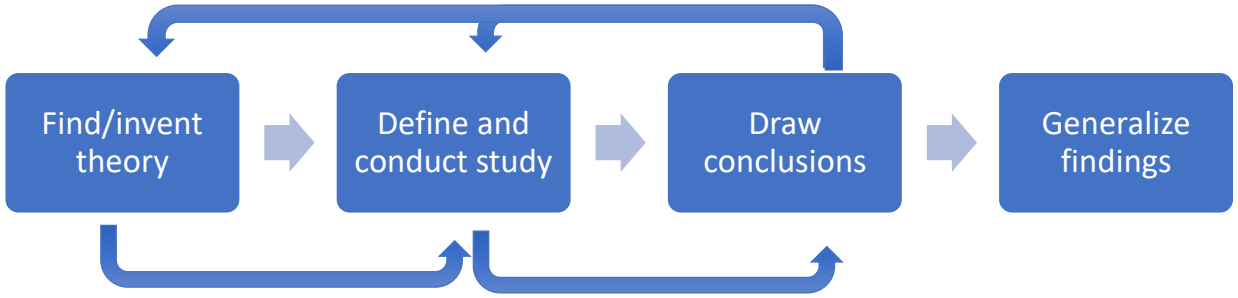
Controlled Experiment vs. Field Study

- Control influence of confounding factors in controlled lab experiments
  - One sorting algorithm is faster than another on the data X on computer Y
  - Reliable measurement of dependent variable (i.e., high internal validity)
  - Results cannot be generalized to other circumstances (i.e., low external validity)
  - Not always possible, based on practical and ethical reasons

- In the field, confounding factors cannot always be controlled for
  - Low internal validity
  - Higher external validity

Field Studies

- Interpretation of verbal material
- Focus on experience
- Open questions
- „More details than one measurement value“
- Realistic settings instead of lab settings
- Controlled experiments are strict in their execution, qualitative studies have flexibility and are more iterative



- No statistical significance tests
- Difficult to compare

Qualitative and Quantitative Methods

- Often, both methods are combined

- Using a new UI:
  - Observation of users while they use new UI
  - Observe strategies of users and abstract
  - Measure time users need to complete a task
  - Evaluate relationship between strategy and time

- Our neuroimaging studies often include a post-session interview

Interviews

Interviews

Interviews are a qualitative research method that collects subjective views by asking questions.

- Interviews can involve two or more people.
- Participants are often happy to participate and share their views
- Unlike other research methods, it allows for direct clarification
- But it is less reliable (participants can be biased in many ways), sampling is critical, and it is time-consuming
- For example: we would like to understand how programmers see testers and vice versa with the goal to develop better collaboration strategies
  - → we interview programmers and testers at a local company

Interviews

Interviews can be differentiated among several features:

- Degree of standardization (structured, semi-structured, unstructured)
- Number of interviewees (single interview, group interview, survey)
- Number of interviewer (one interviewer, tandem, hearing)
- Kind of contact (direct, via phone, via mail)
- Authority of interviewer (soft, neutral, hard)

Structured Interviews

- Questions and order is clearly defined before starting the interview
- Mostly closed questions
  - “Do you prefer dynamically or statically typed programming languages?” → Binary response
- Possible answers are prepared and only need to be marked

- Suitable for well-known topics in which you can formulate clear questions
- Structured interviews tend to be short and efficient to collect as well as to analyze

Semi-Structured Interviews

- Pre-defined questions with a rough outline of the discussion
  - "Do you prefer dynamically or statically typed programming languages?" → Binary response
  - "If dynamically/statically, can you state one reason for it?"
- But, open questions and answers allowing flexibility
- In many cases in SE, it is well suited for exploratory work
- Example: post-interviews for our neuroimaging studies

Unstructured/Open Interviews

- Research and field interviews
- Open questions
  - “What do you think about typing in programming languages?”
- More like a conversation than a question/answer situation
- Especially suitable for exploration of novel topics to gather data for formulizing hypotheses
- Interviewer must not influence interviewee and must not show own opinion (which requires a lot of experience)

Interviews: Comparison

Structured	Semi-Structured	Unstructured
Can result in quantitative data	Can result in quantitative data	No quantitative data
Makes comparisons possible	Some comparisons possible	Comparison difficult
High reliability	Aspects of high reliability	Low reliability
Limitations in scope and questions	Additional questions allow for larger scope	Open scope
Biasing responses through closed questions	Biasing responses through closed questions	Biasing responses
Fairly time efficient	Less time efficient	Can take a lot of time (depends on interviewee)
	Requires skill from the interviewer to ask the right additional questions	Requires skill from the interviewer to guide the entire conversation

Interviews: Group Interviews

- It is possible to interview multiple people at once (“focus group”)
- The interviewer acts more as a moderator and guides an open conversation between participants
  - Can be easier on the interviewer as it is not a forced one-on-one setting
- Typically very open (“trigger”) questions to explore a new topic
  - Looking for ideas and directions for future research
- More time efficient to get the viewpoint from multiple people
- But can become unfocused and is harder to schedule

Interviews: Formulating Questions

- Order of the questions is critical
  - Start with generic demographic questions (easy to answer, eases participants into the format)
  - If applicable, ask for specific demographics (position in company, …)
  - Then, general questions on your topic
  - Then, go into more detail, but keep topics separated
  - Ideally, the questions create a flow in the interview
- Use language that is understandable to the interviewee
- Do not ask leading questions

Interviews: Organization

1. Content Preparation
   - Why? What are your goals?
   - Define topics and participants
   - Study domain knowledge/terms
   - Prepare specific questions
     - Pilot and iterate over questions
   - Write down guideline for the interview
2. Organizational Preparation
   - Identify and recruit interviewees
   - Setup devices (audio recording, zoom call, …)
   - If collaborating with others, instruct them and discuss approach in advance
3. Interview
   - Before arrival, check technical devices (disk space?)
   - Get to know the interviewee (a bit)
   - Explain process and obtain informed consent
   - Start by asking demographic questions (easy to answer, ease participants into the interview)
   - Conduct interview
   - Conduct post-interview (explain goals, ask for further comments, …)
   - Stop recording
   - Write down notes and save/backup recording

Interviews: Successful Interviewer

A good interviewer shows several important characteristics:

- **Knowledgeable**: is familiar with the focus of the interview
- **Structuring**: provides purpose for the interview, inquires for questions from the interviewee
- **Clear**: asks simple, easy, short questions
- **Gentle**: lets interviewees finish, gives time to think, tolerates pauses
- **Sensitive**: listens attentively to what is being said and how it is said, is empathic in dealing with interviewee
- **Open**: responds to what is important to interviewee and is flexible
- **Steering**: knows what they would like to find out
- **Critical**: is prepared to challenge what is said (e.g., dealing with inconsistencies in interviewee's replies)
- **Remembering**: relates what is said to what has been previously said
- **Interpreting**: clarifies and extends meaning of interviewee's statements, but without imposing meaning onto them
- **Balanced**: does not talk too much (and not too little)
- **Ethically sensitive**: is sensitive to the ethical dimensions of interviewing, ensuring the interview appreciates what the research is about, its purposes, and that their answers will be treated confidentially
- **Adaptable**: each interviewee is different and will have to adapt the behavior accordingly

Interviewing is difficult and requires experience to do well!

Interviews: Your First Interview

If you are conducting your first interview, consider the following challenges

- Prepare for the unexpected interviewee behavior or environmental problems:
  - Prepare for unexpected answers. What do you do with brutally honest answers? How can you guide the conversation back?
  - Expect noise in the environment, interview interruptions (e.g., Internet problems)

- Intrusion of own biases and expectations
  - Be careful to not ask leading questions and not influence the interviewee
  - Practice prepared questions and ideally follow a script

Interviews: Your First Interview

If you are conducting your first interview, consider the following challenges

- Maintain focus
  - Pass to the next question only when you are satisfied with the answer to the current question → otherwise ask probing/clarifying questions
  - Do not hurry, this is your only chance to get the information
  - Do not schedule too many interviews in a day

- Dealing with sensitive issues
  - Sometimes interviewees get uncomfortable with some questions, be receptive and change topic

Interviews: Transcription

- Transcription
  - Requires high effort
  - About one page of text per minute
- Archiving of material
  - 10 years (according to German Research Foundation)
- Data privacy
  - Pseudonymize/Anonymize data
  - Destroy or return raw material

Interviews: Transcription

- Transcriptions can be extremely detailed

Box 5.2

Einige Transkriptionszeichen

Transkiptionszeichen	Bedeutung
montag kam er ins krankenhaus	Interviewtext (nur Kleinschreibung!)
MONtag kam er ins krankenhaus	Betonung von Silben durch Großschreibung
MONtag kam er * ins krankenhaus	Kurzpause durch *
MONtag kam er ** ins krankenhaus	längere Pause durch **
MONtag kam er *2* ins krankenhaus	Pause über 1 Sek. mit Längenangabe *Sek.*
MONtag kam er *2* ins kranken/	Abbruch eines Wortes oder Satzes durch /
MONtag kam er *2* in=s kranken/	Wortverschmelzung durch =
MONtag kam er *2* in=s krank'n/	ausgefallene Buchstaben durch '
MONtag kaaam er *2* in=s krank'n/	Dehnung durch Buchstabenwiederholung
MONtag kaaam er *2* in=s krank'n/ (WEINEN)	Kommentar in Klammern und Großbuchstaben
MONtag kaaam er *2* in=s <krank'n/ (WEINEN)	Tonhöhe fallend < (steigend: >)
I: #Wann# A: #MONtag# kaaam er *2* in=s <krank'n/ (WEINEN)	gleichzeitiges Reden von Interviewer (I) und Befragungsperson (hier: A) markiert durch Doppelkreuz (#)

Interviews: Potential Issues

In summary, interviews can be challenging, because…

- It is tricky to be a good (likeable) interviewer that gets the most out of people
- Identifying the right people is tricky, and confidentiality may prevent them from disclosing important, relevant information
- Many interesting participants may not have or want to take the time
- Technical people will speak technical language
- Interviews produce a lot of data, and it takes a lot of effort to transcribe and analyze

Yet, interviews are the best tool to get to know people and their knowledge.

Case Studies

Case Study

- Detailed observation of a single object or of a few selected examples
- Many applications, but in SE often in UI research
- Gain insights into a specific context-dependent topic in a real-world scenario
- Often focus on a specific small sample (even outliers)
- Examples:
  - Observing how developers handle a new tool
  - Applying a new programming paradigm to an existing implementation

Case Study to Build Theory

- Pilot study or an explorative study to provide new/unexpected insights
- In early phases of a project to guide future directions
  - Sometimes opens new research directions
- To build theories, which can then be evaluated quantitatively
  - Strengthen an existing theory by supporting the investigated case
  - Expand an existing theory by uncovering new elements
  - Question an existing theory by showing a negative case

Case Study and Quantitative Measurement

- Quantitative measurement within a case study is possible
  - For example, increased speed in completing a task with new UI
  - Increased speed with a new database index

- No conclusion for general cases (low external validity)

Typical Criticisms of Case Studies

- Uncontrolled and subjective → not reliable
- Tendency to confirm existing hypotheses
- Not generalizable
- Many details, difficult to summarize

Experience through Case Studies

- Observe a problem within a context
- Learn from individual cases
- Realistic details
- No abstraction/simplification
- No hard data, but valuable experiences

Case Studies for Falsifying

- Case study can falsify a hypothesis
- One well-selected example can suffice

- Example:
  - Galileo's experiment for gravity (feather vs. lead) with case study instead of series of experiments

Case Selection

Selection	Rationalization
Random	Reduces bias; may be generalizable
Extreme case	Unusual case, especially problematic or especially suitable; makes one point pretty clear
Maximal variation	Several very different cases (e.g., three cases that differ in size/language/experience)
Critical Case	Allows to make conclusion like: „When it is (not) working here, it (does not) work(s) in other cases“; e.g., to assess plausibility of a theory
Paradigmatic	General typical case, which is used by several researchers; theories may be based on this case

Further Reading

- Studying Software Engineers: Data Collection Techniques for Software Field Studies (https://link.springer.com/content/pdf/10.1007/s10664-005-1290-x.pdf)
- Five Misunderstandings About Case Study Research

(https://arxiv.org/ftp/arxiv/papers/1304/1304.1186.pdf)

Questionnaires

Questionnaires

- Questionnaires allow systematically collecting qualitative and quantitative data through predefined questions
  - They can be sent to many people at once (“Surveys“)
  - Typical examples are customer satisfaction, political polls, usability of a tool, …
- Similar to an experiment or an interview, there often is a specific target population
  - For example, Python programmers, University students in computer science, …
- Questionnaires appear to be trivially easy, but “the devil is in the details”
- Often used in SE practice and research, but comparatively superficial
  - They are usually easy to answer
  - But it is nowadays difficult to get responses (<10% response rate) due to survey fatigue, which leads to further biases

Questionnaire

- How would you design a questionnaire for „How good is this empirical software engineering research course?“
- What kind of question(s) would you ask?

Questionnaires: Process

- Define goals of the questionnaire
- Define and understand participant sample
- Implement questionnaire (electronically, paper, online?)
- Set up and test questions and answers
- Collect data
- Analyze data
- Report

Questionnaires: Recruiting

While questionnaires can be sent around freely, in many cases a targeted approach is sensible

- Send individualized email invitations (rather than a generic “Dear Sir or Madam”)
  - Increases likelihood of being read and participation
- Establish an official end of the survey (on March 18th 2023, not too far out!)
- Offer incentives (raffles, donations, direct payment, sharing results)
- Send questionnaires with individual tokens or tracking codes for specific broadcasting avenues

Questionnaires: Setup

Questionnaires typically do not just contain questions, but some “wrap” around it

- An explanation of the purpose of the study
- A realistic estimate of time required to complete the questionnaire
  - Do not use an unrealistically low estimate, it will be counterproductive
- A description who is conducting the study
- A contact name and email address
- If applicable, ethical review board approval, privacy and data protection statements
- Sometimes also an explanation of how the respondents were chosen and why

Questionnaires: Types of Questions

- Demographic questions: *How old are you?*
- Personal factual questions: W*hat is your role in the company/organization?*
- Factual questions about others: *How old are, on average, your colleagues?*
- Informant questions: *Does your organization employ external suppliers?*
- Questions about attitudes: *Is your job interesting?*
- Questions about believes: *Does bad code lead to more errors?*
- Questions about standards and values: *Is it appropriate to wear shorts at your organization?*
- Questions about knowledge (rare): *What are three requirements for good identifier names?*

Questionnaires: Questions and Answers

- Ask closed questions and obtain quantitative answers
- For example, through a Likert scale, often 1-5 or 1-7
  - Example from our research: How experienced are you with the following programming languages?

	Very inexperienced	Inexperienced	Neither /nor	Experienced	Very experienced
Java	1	2	3	~~4~~	5
C	1	2	~~3~~	4	5
Haskell	1	2	~~3~~	4	5
Prolog	1	~~2~~	3	4	5

Likert Scales

- Likert scales are a widely used rating scale to measure opinion/attitudes
  - They provide more nuanced insights than yes/no responses
- They can be one directional („Never“, „Sometimes“, „Always“)
- They can be bi-directional („Easy“, „Neutral“, „Difficult“)
- Quick and easy for participants, but forces them into a specific range
- Helpful when there are no possible quantitative measurements
  - „How experienced are you with Java?“

Likert Scales

- Likert scales usually provide ordinal data
  - Inferential statistics: chi-square test, …
- Some argue in some cases it is interval data, but it is difficult to show the distance between the response options is the same
  - Is “Strong agree” twice as agreeable as “Agree”?
  - If so, you can use ANOVA, …

Likert Scales

- Response/social desirability bias: participants tend to shy away from extreme opinions (that may be perceived negatively)
- Can be boring for participants leading to inattention or clicking in some pattern (left/right-left/right, …)
  - "Just quickly click through it"
  - Some researchers use "trick" questions ("I am still paying attention: Agree/Neutral/Disagree" in the middle of the questionnaire to check for attention

Questionnaire: Formulating Questions

- Be very cautious when phrasing questions.
  - Unlike an interview, questionnaires will not reveal misunderstandings
- Avoid vague or ambiguous questions and answer pairs
  - ✗ „How often does your group have meetings“: Never/Sometimes/Often
  - ✓ „How frequently does your group have meetings“: Multiple times per day/Once a day/…
- They must be clear and avoid double-negatives
  - ✗ „I never write code in a non-IDE-environment!“: Agree/Neutral/Disagree
  - ✓ „I only write code in an IDE-environment!“: Agree/Neutral/Disagree
  - ✓ „I prefer writing code in a text editor“: Agree/Neutral/Disagree

Questionnaire: Formulating Questions

- Inquire about one thing at a time
  - ✗ „I prefer writing code in an IDE or in a text editor“: Agree/Neutral/Disagree
  - ✓ „I prefer writing code in an IDE“: Agree/Neutral/Disagree
- Avoid long questions
  - ✗ „What types of defects are typically encountered by programmers whose relevance is normally difficult to communicate to team leads?
- Avoid generic questions
  - ✗ „What is the physical, intellectual, and moral condition in your group?“
- In general, forced choice answers (“Do you use X: yes/no”) are more reliable than “Do you use X: check all that apply”

Questionnaire: Closed Questions

- Questionnaires can include closed questions to gather specific characteristics
- Typically binary, nominal, or ordinal data
- "What is your current occupation?"
  - Student
  - PhD-Student
  - Professor
  - Other
- "What is your age?"
  - Under 18
  - 18-34
  - 35-65
  - Over 65

Questionnaire: Open Questions

- Open questions allow to gather subjective views (open text field)
- Allows participants to respond in unknown/unexpected ways
- Typically
  - The responses are shorter than during an interview
  - Despite asking complex questions, participants often are very brief as it takes time
- Open questions must be of high quality, clear, one topic at a time, …
  - Avoid biasing questions:
    - “All SE experts agree Python is an excellent language. What do you think?”
    - “Empirical software engineering research is an excellent course, right?”

Questionnaires: Incorrect/Nonsensical Responses

- As discussed in the data analysis lecture, prevent as much as possible with validation rules

Question	Response
Immatriculation	1945
Since how many years are programming?	99
How many programming classes did you attend?	1945

Surveys: Tasks

- In SE, “surveys“ can include some tasks
  - For example, ask programming experience and demographic questions
  - Then solve several program comprehension problems
- Allows for easy participation and collecting large amounts of data
  - But, often lower quality than lab experiments
  - Chances of participants not taking it as serious or getting distracted

Design

- Read literature
- Ask experts
  - If possible, use existing questionnaires
  - In psychology, you cannot use a new questionnaire without testing it in its own study first
- Consider ordering effects
  - We ask for self-estimation of programming skills **before** an experiment to avoid biasing

Questionnaires: Advantages

- Cheap
- Easy/Easier to reach large samples
- Well-suited to complement other methods
- Doable online (but potential for confusion, distraction, …)
- But, limit the length and motivate participants

- Tools: Soscisurvey, SurveyMonkey, EFSSurvey, …

Other Qualitative Studies

Other Qualitative Studies

- Conceptional modelling: participants draw a diagram of some aspect of their work/life (e.g., organizational structure, workflows) often in combination with an interview
  - Turn implicit mental models into explicit data
  - Easy to collect and rich data structure (in comparison to verbal transcripts)
  - But require domain knowledge to understand and interpret
- Work diaries: participants record events during their workday
  - More accurate than self-reflection in retrospect
  - But hard to convince people to do it in high quality as it interferes with their daily work

Other Qualitative Studies

- Think-aloud protocols: participants verbalize their thoughts during a task
  - Turn implicit mental models into understandable insights
  - Easy to collect, but difficult for the participants (especially for cognitively demanding tasks)

- Individual observation: researchers “shadows” a participant throughout their activity/work
  - Easy to implement but time-consuming for the researcher
  - Can be uncomfortable for the participants (→ risk of bias)
  - Superficial insights only on a directly observable level

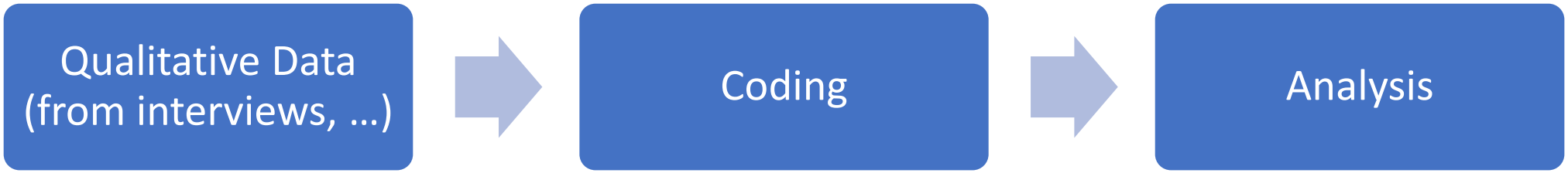
Other Qualitative Studies

- Group observation (“fly on the wall”): video recording of group activities, such as meetings
  - Little effort for the researcher and no effort for the participant
  - But difficult to analyze as video contains many layers

Data Analysis of Qualitative Data

Data Analysis of Qualitative Data

- Text:
  - Structured (e.g., closed, semi-structured interviews/questions)
  - Unstructured (e.g., open, semi-structured interviews/questions)
- Audio recording (e.g., think-aloud protocols)
- Video recording (e.g., screen recordings, group behavior, …)



Data Analysis: Steps

1. Record and document data
   - If applicable, make transcripts
2. Classify/encode data in clusters/categories/concepts
3. Connect data to understand relationships between clusters/categories/concepts
4. Validate data by evaluating alternative explanations and negative cases
5. Report

Data Analysis

- Inductive approach:
  - No prior knowledge
  - Build concepts/categories while analyzing data
  - Iterative approach

- Deductive approach
  - Prior knowledge exists, for example with research questions
  - Grouping of the data based on prior knowledge

Data Analysis: Coding

- A code is a brief word/phrase to summarize/represent some aspect of the data
- Codes are organized in a coding framework that is setup in a hierarchy
- Coding often requires multiple iterations, especially if the coding framework is enhanced on the fly
- Typically used for interview transcripts, but can be used for all kinds of data
  - For example, tagging images or videos with codes

Data Analysis: Coding

- Coding is not just labeling text, it is abstracting and understanding and therefore can be very subjective
  - If possible, use multiple researchers independent of each other for coding and compare their agreement
  - Cohen's kappa for inter-rater agreement
  - If low: discuss, ask additional researchers, …
  - Report Cohen's kappa before and after discussion

Kappa	Strength
< 0.20	Very weak
0.21 – 0.4	Weak
0.41 – 0.6	Medium
0.61 – 0.8	Strong
0.81 – 1	Very strong

- Coding is time intensive (for example, full day for 90min of transcripts)

Data Analysis: Content Analysis

- Content analysis aims at identifying patterns in text data
- It can be seen as reliable and replicable technique to transform large amounts of data into small, structured concepts
- “A priori coding” or closed coding considers already set up coding framework before collecting data (based on hypotheses, theory, …)
- “Emergent coding” or open coding corresponds to creating coding while analyzing the data
  - “Grounded Theory” as a systematic approach to construct a new theory

Data Analysis: Content & Statistical Analysis

- After content analysis, statistical analysis can be possible
  - For example, experts mention “testing” more often than novices (in a statistically significant way, chi-square …)

Further Reading

- Lecture on Coding, Thematic Analysis, Grounded Theory:
  https://www.youtube.com/watch?v=UAfbkIWX7jc