Using Qualitative Methods in Empirical Studies of Software Engineering

Carolyn Seaman
University of Maryland Baltimore County
Fraunhofer USA Center for Empirical Software Engineering Maryland
Universidade Federal de Pernambuco

USP
26 March 2013
São Paulo, Brasil
Outline

- What, when, why qualitative methods?
- Data collection techniques
  - Participant observation
  - Interviewing
    - Hands on exercise
- Data analysis techniques
  - Coding
  - Constant comparison method
    - Hands on exercise
- Verification
- Mixed methods

© Carolyn Seaman, 2013
Definitions

- **Qualitative data** - data in the form of text and pictures, not numbers
- **Qualitative analysis** – analysis of qualitative data in order to discover trends, patterns, and generalizations
- **Grounded theory** – theory formed bottom-up from the (usually qualitative) data
- **Rich data** – data that includes a lot of explanatory and context information
Why Qualitative Methods?

- **Problem**: Difficult to answer complex SE questions with a purely quantitative approach because
  - Working with human subjects
  - Typically have small sample sizes
  - Experiments are expensive to run
  - Need some support for a hypothesis before investing effort in full experiment

- **Solution**: Use a qualitative approach that includes a quantitative aspect
Types of results

A qualitative study will result in:

- Propositions tied to a trail of “evidence”
- Well-grounded hypotheses
- Complex findings that incorporate the messiness of the phenomenon under study
- Explanations
- Areas for future study
Types of Research Questions

Qualitative methods are most appropriate when:

- Subject of study involves human behavior
- No concrete hypotheses
- Variables hard to define or quantify
- Little previous work
- Quantitative results may be hard to interpret

© Carolyn Seaman, 2013
Advantages to Researchers

- Richer results
- Results more explanatory
- Closer to sources of data
- Avoid errors in interpretation
Advantages to Practitioners

- *Richer*, more relevant results
- *Terminology* of results
- More part of the research process
- Opportunity to clarify and explain findings
Overview of Techniques

Data Collection

- Prior Ethnography
- Participant Observation
- Interviewing
- Surveys
- Document Analysis

Data Analysis

- Coding
- Constant Comparison Method
- Cross-case analysis
- Member checking
- Auditing

© Carolyn Seaman, 2013
Participant Observation

**Definition**: non-covert direct observation of phenomenon

**Example**: Observation of code inspection meetings

- collected both qualitative and quantitative data
- did not participate in the inspection
- used data forms as well as field notes
# Observation Data Form

**Inspection Data Form**

<table>
<thead>
<tr>
<th>Class(es) inspected</th>
<th>Inspection date:</th>
<th>Time:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Author:**

**Moderator:**

**Reviewers:**

<table>
<thead>
<tr>
<th>Name</th>
<th>Responsibility</th>
<th>Preparation time</th>
<th>Present</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Amount of code inspected:**

**Complexity of classes:**

**Discussion codes:**

- **D** = Defects
- **Q** = Questions
- **C** = Classgen defect
- **U** = Unresolved issues
- **G/D** = Global defects
- **G/Q** = Global questions
- **P** = Process issues
- **A** = Administrative issues
- **M** = Miscellaneous discussion

**Time logged (in minutes):**

D______ Q______ C_____ U_____ G/D______ G/Q______ P_____ A_____ M______
Field Notes Example

The "step" function is a very important but complicated function. [Reviewer1] did not have time to review it in detail, but [Author] said he really wanted someone to go over it carefully, so [Reviewer1] said she would later.

There was a 4-minute discussion of testing for proper default values. This is a problem because often the code is such that there is no way to tell what a particular variable was initialized to. [Reviewer2] said "I have no way to see initial value". This was a global discussion, relevant to many classes, including [Reviewer2]'s evidently.
Interviewing

- Interviews are good for getting
  - opinions
  - feelings
  - goals
  - procedures (both formal and informal)
- not facts
Standard Interview Formats

- **Structured** (standardized)
  - Tightly scripted, almost verbal questionnaire
  - Replicable, but lacks richness
  - Analyze like questionnaire
  - “How many times a day do you access the internet? [0, 1-5, 5-10, 10-15, 15+]”
Standard Interview Formats

- **Unstructured**
  (Open/Informal/Conversational)
  - Guided by a very scant script.
  - Rich, but not replicable.
  - Difficult to be systematic, problem of coverage.
  - Minimize interviewer effects, preserves interviewee point of view.
  - Interviewee led, interviewer probes.
  - “Please, tell me about your internet usage...”
Standard Interview Formats

- **Semi-structured**
  - Guided by a script (interview guide), but interesting issues can be explored in more depth.
  - Good balance between richness and replicability.
  - Mixed analysis techniques.
  - “In a typical day, how often do you use the internet?”
Interview questions

- Closed
  - Predetermined answer format (e.g. Yes/No)
  - Easier to analyze
- Open
  - No predetermined answer format
  - More complete response
- Combination
  - Closed, with opportunity to elaborate
- Probes
- Pitfalls:
  - leading questions
  - double-barreled questions
  - judgmental questions
Interview Guide

- A script for use by interviewer only
- “Wish list” vs. structured
- Flow/direction to interview
- Required topics
- Transitions between topic areas
- Important for replicability
- Wording and sequence are critical
Interview Design Considerations

- Context switching
- Flow between open and closed questions
- “Shape” of interview
- Most important stuff first
- Wording
Interview Shapes

- **Funnel**
  - Begin with open, gradually become more closed
  - Good if you’re not sure what you’re going to get

- **Pyramid**
  - Begin with closed, gradually become more open
  - Good with nervous interviewees

- **Hour glass**
  - Begin with open, gradually become more closed, then open up again at end to pick up things you might have missed
  - Good if you know what you want, but suspect there are important things you don’t know about yet
Interviewing Pointers

- give clues about the level of detail you want
- establish rapport, but be subject neutral
- avoid jargon, esp. academese
- dispel any notion of the “right” answer
- play the novice when appropriate
- probe, but do not lead
- always be aware of your biases
- be sensitive to their work (environment/schedule)
- no more than 60 minutes
- let interviewee know next steps
- end with “anything else I should know?”
- say Thank you!
Recording of interviews

- Audiorecording
- Notetaking
- Scribing
Audiorecording

- Best memory mechanism
- Full transcription or just verbatim quotes
- Still take notes
  - Tapes fail, digital files are deleted
  - Does not record all aspects (esp. context / facial expressions)
- Required consent
  - Always ask first.
  - Do **NOT** hide recorder, keep it visible at all times.
  - Give the option to turn it off at any point.
Notetaking

- Very hard to take notes and interview at the same time
- There are some super-researchers who can do it
- Inevitably results in incomplete notes
- Slows down the interview
- Sometimes inevitable
Scribing

- Partner-based interviewing
- Advantages of a single contact vs. trading-off
- Can share roles (interviewer/scribe)
  - BOTH take notes, though to different degree
- Group debrief: what did you get/miss?
- Synchronize notes: overlap and emphasis
- Clarify while it is still in your head
Writing up the interview

ASAP!!!!
Interview Notes

- Write it up immediately
- Descriptive vs. reflective notes
- Use Observer’s Comments
  - Impressions, state of mind, assumptions, notes to self
- How detailed?
  - Verbatim transcript
    - only possible with audiorecording
    - Extremely labor-intensive
  - Summaries with major points quoted
    - OK, but use LOTS of quotes
  - Start closer to verbatim at the beginning of a study
Interviewing Exercise

- **Background:**
  - The National Federation of Makers of Feijoada (FNFF) is concerned that the national consumption of feijoada is declining due to decreasing quality of feijoada.
  - So they have asked us to interview the top feijoada chefs in the country (as determined by regional competitions).
  - The goal is to find out the secrets to master feijoada making, so that it can start to be taught in elementary schools.
Interviewing Exercise

- Three versions of the interview guide
- I will be the interviewer
- You will be the interviewees
  - So take a moment to think of your favorite feijoada recipe
Interviewing Exercise

1. Recap

First interview: pyramid
- Started with easy, closed questions
- Ended with open-ended questions

Second interview: funnel
- Started very broadly, with open questions
- Followed up with narrower, closed questions

Third interview: just bad
- Leading, judgmental questions
- Double-barreled questions
- Switching from topic to topic
- Switching between open and closed

1. What do you think makes your feijoada the best?
2. How often do you make feijoada and how long does it take you?
3. Of course, you always wash your hands thoroughly before you start, right?
4. Do you add the sausage near the beginning or near the end of the cooking?
5. What kind of pot do you use?

1. How often do you make feijoada and how long does it take you?
2. What do you think makes your feijoada the best?
3. Of course, you always wash your hands thoroughly before you start, right?
4. Do you add the sausage near the beginning or near the end of the cooking?
5. What kind of pot do you use?
Constant Comparison Method

- Qualitative analysis method
- Meant to generate grounded theory
- Operates on a set of field notes
- Basic process:
  - coding
  - grouping
  - writing field memo
  - forming hypotheses
- Repeated periodically in parallel with data collection
What’s a Code?

- A label
- A concept
- A topic
- A category
- A relationship
- A theme
What's Coding?

- Open coding - assigning codes to pieces of textual data
  - Coded “chunks” can overlap
  - One chunk can have several codes
- Axial coding - grouping, categorizing, combining coded chunks
- Selective coding - making sense of it
Open Coding

What’s here? What are the pieces?

- Identification/discovery of concepts
- Classification (labeling of phenomena)
- Abstraction (this is part of that)
- Comparative analysis (this is different from that)
- Categorization (organization, grouping)
- Value-neutral, at least initially
  - “complexity” not “high complexity” or “low complexity”
Open Coding Process

- Preparing for coding
  - Read the data
  - Read background material and research design
  - Create pre-formed codes, if applicable
- Coding by hand
  - Document markup (colored pens, etc.)
  - Photocopy, scissors, and envelopes
  - MS Word comments
  - Excel
- Coding tools – NVivo, Atlas TI
- Coding scheme
  - *Pre formed* or *post formed* codes
  - Constant iteration
  - Structure develops over time
Open Coding Exercise

- **Background:**
  - Study sources of information in software maintenance
  - Interviews with experienced software maintainers in several organizations

- **Process:**
  - I’ll show you an example
  - Then you’ll try it – code one excerpt with one code
  - Find a partner – compare your codings
  - I’ll show you my coding of the excerpt
Coding Scheme

Respondent Background
Information Gathering
Transition to maintenance
Types of documentation
Characteristics of Documentation
Quality of documentation
Properties of documentation
Missing documentation
Creating documentation
Location of documentation
Importance of documentation

Human Sources of Information

Quality of Process

Great Quotes

© Carolyn Seaman, 2013
Open Coding and Quantification

- One form of coding
- Objective is to derive quantitative data from qualitative data for future statistical analysis
- Usually involves counting
  - How many subjects said…?
  - How many times did subjects use the term …?
  - How many times did …?
- Or timing
  - How long did subjects spend doing…?
  - How long did it take to …?
- Inevitably loses richness
- Often seems a little like missing the point
  - What’s the point of collecting rich data when you’re just going to condense it down to numbers?
- But often is an effective and necessary way to reduce the size of the data
## Inspection Data Form

<table>
<thead>
<tr>
<th>Class(es) inspected:</th>
<th>Inspection date:</th>
<th>Time:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderator:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewers:</td>
<td>Responsibility</td>
<td>Preparation time</td>
</tr>
<tr>
<td>Name</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Amount of code inspected:**

**Complexity of classes:**

**Discussion codes:**

**D** Defects  
Reviewer raises a question or concern and it is determined that it is a defect which the author must fix; time recorded may include discussion of the solution

**Q** Questions  
Reviewer asks a question, but it is not determined to be a defect.

**C** Classgen defect  
Reviewer raises a defect caused by classgen; author must fix it, but it is recognized as a problem to eventually be solved by classgen.

**U** Unresolved issues  
Discussion of an issue which cannot be resolved; someone else not at the meeting must be consulted (put name of person to be consulted in () beside the code); this includes unresolved classgen issues. It also includes issues which the author has to investigate more before resolving.

**G/D** Global defects  
Discussion of global issues, e.g. standard practices, checking for null pointers, which results in a defect being logged (does not include classgen defects)

**G/Q** Global questions  
Same as above, but no defect is logged

**P** Process issues  
General discussion and questions about the inspection process itself, including how to fill out forms, the order to consider material in, etc., but not the actual execution of these tasks.

**A** Administrative issues  
Includes recording prep time, arranging rework, announcing which products are being inspected, silence while people look through their printouts, filling out forms.

**M** Miscellaneous discussion

**Time logged (in minutes):**

D _____ Q _____ C _____ U _____ G/D _____ G/Q _____ P _____ A _____ M _____
<table>
<thead>
<tr>
<th>Time</th>
<th>Participants</th>
<th>Code</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>SM</td>
<td>A</td>
<td>get started, make a new file, find right files</td>
</tr>
<tr>
<td>31</td>
<td>AP, PK</td>
<td>G/D</td>
<td>change to null</td>
</tr>
<tr>
<td>32</td>
<td>AP, SM, PK</td>
<td>G/D</td>
<td>actually several small fixes needed</td>
</tr>
<tr>
<td>34</td>
<td>SM</td>
<td>G/D</td>
<td>don't change now, wait for next week</td>
</tr>
<tr>
<td>35</td>
<td>M1</td>
<td>G/D</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>M1</td>
<td>G/D</td>
<td>only through fifth to eighth, not all, do two more for next week</td>
</tr>
<tr>
<td>38</td>
<td>M1, PK, SM</td>
<td>G/D</td>
<td>M1 went through everything, still need to check on pages, don't change for now, re-read please</td>
</tr>
<tr>
<td>40</td>
<td>PK, SM</td>
<td>G/D</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>SM</td>
<td>A</td>
<td>nothing in category</td>
</tr>
<tr>
<td>42</td>
<td>SM, PK</td>
<td>G/D</td>
<td>null instead of 0, had trouble finding it</td>
</tr>
<tr>
<td>44</td>
<td>SM, PK, M1</td>
<td>D</td>
<td>Parameter Error exception - trying to figure out where it's broken</td>
</tr>
<tr>
<td>46</td>
<td>SM, PK, M1</td>
<td>L</td>
<td>similar to above</td>
</tr>
<tr>
<td>47</td>
<td>SM, PK, M1</td>
<td>L (or L)</td>
<td>Rx catching error that will never happen, you're breaking it a lot more complex than you thought it - be careful when checking</td>
</tr>
<tr>
<td>53</td>
<td>PK, SM</td>
<td>Q</td>
<td>clarification</td>
</tr>
<tr>
<td>55</td>
<td>PK, M1, SM</td>
<td>Q</td>
<td>Parameter Error - handle differently from the way classroom does it</td>
</tr>
<tr>
<td>56</td>
<td>SM, PK, M1</td>
<td>D</td>
<td></td>
</tr>
</tbody>
</table>

Lots of time for everyone trying to find right place in printout - small print is a factor
Axial Coding

How are things related?

- Initial process of reassembling
- Relationships among categories and codes
- Structure (why?)
- Process (how?)
- Explanations not causal prediction
Selective Coding

How does it all fit together?

- Also called **sense making**
- Relationships among relationships
- Theory construction
- The central category
- Storyline memos
- Role of literature
- Write, write, write!!!
- **Field Memos**
Field Memos

- The “single most powerful analytical tool” for qualitative researchers
- Simply, a piece of writing
- Maybe will later become part of a report, maybe will be thrown out
- Summarizes and synthesizes:
  - A proposition
  - An open question
  - A chain of evidence and logic
  - The complexity of a concept
  - Rich description
- Version control and organization
Verification

- Process of establishing a study’s trustworthiness and quality
- Analogous to assessing validity in quantitative studies
  - Relevant quantitative validity issues include internal, external, and construct validity, reliability, etc.
  - Some qualitative researchers simply adopt this terminology but translate
- **Big difference:** in qualitative work, verification is a continuous process that occurs throughout the study
  - Thus verification is an integral part of the techniques used to carry out a study, not a set of techniques applied after the study.
- Multiple sets of terms and concepts exist for verification of qualitative studies
Lincoln & Guba’s Verification Terms

- **Credibility**
  - Length of time and degree of contact
  - Triangulation
- **Transferability**
  - Thick description, lots of context
- **Dependability**
  - Results not subject to change and instability
- **Confirmability**
  - Strength of chain of evidence
Verification Techniques

- Prolonged engagement and persistent observation
- Triangulation
- Peer review and debriefing
- Negative case analysis
- Clarifying researcher bias
- Member checks
- Rich, thick description
- External audits
Triangulation

- Simply put, getting your evidence from multiple sources in multiple ways
- Ideally, each proposition put forth should be supported by data that is
  - From at least two different sources,
  - Of at least two different types, and
  - Collected in at least two different ways
- Trick is to merge data during analysis, but keep track of where it came from
Negative Case Analysis

- Search for data that will *disconfirm* your proposition
- If you don’t find it, be able to show *convincingly* that you tried
- If you do find it, show how you *modified* your proposition to reflect it
- Negative evidence *doesn’t mean you’re wrong*, just that you have to bend a little
- Requires *constant skepticism* – sometimes not possible for an immersed researcher – need a *skeptical buddy*
Member Checks

- Checking intermediate propositions, results, findings with *subjects*
- Subjects will *suggest* alternative interpretations, sources for negative evidence, terminology
- A *variety of settings*:
  - Extra round of interviews
  - “Thank you” workshop
  - Wrap-up presentation
  - Sending a report – almost never works
Showing Verification

- Creswell recommends applying \textit{at least 2} of the verification techniques on every study
- I would recommend more
- Transparency
  - Provide \textit{evidence} in your writings that you have applied the techniques
  - \textbf{Examples} of negative cases and how they were handled
  - \textbf{Accounts} of member checks and results
  - Explicitly \textit{describe} data sources and methods to show triangulation
Using Qualitative and Quantitative Methods Together

- Qualitative and quantitative methods best used in combination
- Can simply be used in parallel to address the same research questions
- There are other strategies to better exploit the strengths and weaknesses of the methods

© Carolyn Seaman, 2013
Example Design 1: Statistical Hypothesis Testing with Follow-up Interviews

- Classic design – often done without fully exploiting the interview data
- Example scenario:
  - Blocked subject-project experiment to evaluate a new testing technique
  - Statistical results show that technique is more effective on some applications than on others
  - Qualitative results show why
Example Design 2: Using Grounded Theory to Identify Variables

- Want to evaluate a new technique, but not sure what the evaluation criteria should be

- Example scenario:
  - Evaluating a collaborative design process
  - Use participant observation of design meetings to generate hypotheses about properties of the resulting designs
  - Grounded hypotheses are used to design a quantitative evaluation of the resulting designs

© Carolyn Seaman, 2013
Example Design 3: Using Prior Investigation to Operationalize Variables

- Relevant variables are known, but the range and types of values is difficult to specify
- Example scenario:
  - Want to study the relationship between developer experience and types of defects
  - First use interviews to identify the range of developer experience (in its complexity) and a taxonomy of defect types
  - Quantitative study then is much more effective when using this operationalization
Conclusions

- Empirical software engineering researchers are addressing complex research questions that have human elements
- Qualitative methods, usually in combination with quantitative methods, can be helpful in handling this complexity
- Qualitative methods are both flexible and rigorous
- Qualitative analysis provides richer, more relevant, and more explanatory results
- The most effective research designs combine qualitative and quantitative methods
Bibliography