Empirical Approaches, Questions & Methods

Dewayne E Perry
ENS 623
Perry@ece.utexas.edu

[adapted in part from Steve Easterbrook, U Toronto]
Empirical Approaches

- Three approaches
  - Descriptive
  - Relational
  - Experimental

- Descriptive
  - Goal: careful mapping out a situation in order to describe what is happening
  - Necessary first step in any research
    - Provides the basis or cornerstone
    - Provides the what
  - Rarely sufficient – often what to know why or how
  - But often provides the broad working hypothesis
Empirical Approaches

- **Relational**
  - Need at least two sets of observations so that some phenomenon can be related to each other
  - Two or more variables are measured and related to each other
  - Coordinated observations $\rightarrow$ quantitative degree of correlation
  - Not sufficient to explain why there is a correlation

- **Experimental**
  - Focus on identification of causes, what leads to what
  - Want $X$ is responsible for $Y$, not $X$ is related to $Y$
  - Experimental group versus control group
  - Watch out for problems
Discovery

- **Process of Discovery**
  - **Plausible:** interesting idea
  - **Important:** is it worthy of further consideration?
  - **Acceptable:** do we have a testable theory, can we create an hypothesis for experimental confrontation?
  - **Justifiable:** amenable to evaluation, defense, confirmation?

- **Sources of Discovery**
  - **Intensive case studies**
    - Document certain variables/conditions as prerequisite for a more theoretical study
  - **Paradoxical incidents**
    - Puzzled by contradictory aspects of a situation
  - **Metaphors that stimulate our thinking**
  - **Rules of thumbs, folk wisdom**
  - **Account for conflicting results**
    - Eg, performance in presence of others
Asking Questions

- Asking questions + systematic process to obtain valid answers
  - Make the question clear
  - Hypothesis should be consistent with questions
  - Statement of the problem
  - Critical: asking the right or important questions

- Types of Questions
  - Existence
  - Description/Classification
  - Composition
  - Relationships
  - Descriptive-Comparative
  - Causality
  - Causality-Comparative
  - Causality-Comparison Interactions
Types of Research Questions

- **Existence questions**
  - Does X exist? X is a thing, attribute, phenomenon, behavior, ability, condition, state of affairs etc.
    - Is there a tool that can generate X?
    - Is there a programmer who can write 200k lines per year?
  - Important when controversial
  - Generalization not important, existence is
  - Requires careful scientific work
  - Rule out alternative explanations

- **Description/Classification**
  - What is it like, is it variable or invariant, characteristic limits, unique of member of a known class, a distinctive description?
    - What are the limits of tool X?
    - What are the characteristics of structured programs?
  - Answer requires statements about:
    - Generality and representativeness of sample
    - Uniqueness/distinctness to population
Types of Research Questions

- Composition
  - What are the components of X?
    - What are the principle traits of a good programmer?
    - What are the main factors in a maintainable program?
  - Requires analysis or breakdown of whole into component parts
  - Factor analysis requires care and accuracy
  - Need large enough samples to rule out biases

- Relationships
  - What is the relationship between X and Y?
    - Are exceptions needed for maintainable programs?
    - Is elegance a function of age?
  - For predictiveness, can use multiple regression techniques
  - Or do the relationships fit theoretical models
  - Need valid/reliable measures, sufficient and representative samples, accurate computations, and interpretations supported by the data
Types of Research Questions

- **Descriptive-Comparative**
  - **Is group X different from group Y?**
    - Are Fortran programmers different from Lisp programmers?
    - Do novice C++ programmers make more errors than Java programmers? Experienced programmers?
  - **An elaboration of the simple description question**
  - **Comparison may be organismic**
    - Eg, age, weight, height
  - **Comparison may be socio-economic**
    - Eg, income, job, neighborhood
  - **Must ensure equivalence of other characteristics**
  - **Criteria measures critical - need validity, reliability**
Types of Research Questions

- **Causality**
  - Does X cause, lead to, or prevent changes in Y?
    - Does C++ lead to complex programs?
    - Does using exceptions lead to simpler programs?
  - Manipulate independent variables to get changes in dependent
  - Need control group for non-treatment
  - Must select sample carefully to rule out biases
  - Replications to warrant generality

- **Causality-Comparative**
  - Does X cause more change in Y than Z?
    - Is C++ better than Java in preventing race conditions?
    - Is the Jackson design method better than the Booch method in producing concurrent systems?
  - Compare rival treatments, control
  - Must guarantee that rival treatments are valid and are given in an unbiased manner
Types of Research Questions

- **Causality-Comparison Interactions**
  - Does X cause more changes in Y than Z under certain conditions but not others?
    - Do formal methods work better than informal methods for Europeans but not North Americans?
    - Is the MacOS easier to use than the Windows by naive users but not experienced users?
  - Add more independent variables
Many Methods Available:

- Laboratory Experiments
- Field Studies
- Case Studies
- Pilot Studies
- Rational Reconstructions
- Exemplars
- Surveys
- Artifact/Archive Analysis ("mining")
- Ethnographies
- Action Research
- Simulations
- Benchmarks
Laboratory Experiments

Experimental investigation of a testable hypothesis, in which conditions are set up to isolate the variables of interest ("independent variables") and test how they affect certain measurable outcomes (the "dependent variables")

★ Good for
  ★ Quantitative analysis of benefits of a particular tool/technique
  ★ (demonstrating how scientific we are!)

★ Limitations
  ★ Hard to apply if you cannot simulate the right conditions in the lab
  ★ Limited confidence that the lab setup reflects the real situation
  ★ Ignores contextual factors (e.g. social/org/’al/political factors)
  ★ Extremely time-consuming!

See:
Field Studies

Exploratory study, used where little is currently known about a problem, or where we wish to check that our research goals are grounded in real-life settings; studies organizational practice using anthropological techniques.

✈ Good for
★ Setting a research agenda (what really matters?)
★ Understanding the context for RE problems (naturalistic inquiry)

✈ Limitations
★ Hard to build generalizations (results may be organization specific)
★ Observers' bias

See:

Case Studies

A technique for detailed exploratory investigations, both prospectively and retrospectively, that attempt to understand and explain phenomenon or test theories, using primarily qualitative analysis

❖ **Good for**
  ★ Answering detailed how and why questions
  ★ Gaining deep insights into chains of cause and effect
  ★ Testing theories in complex settings where there is little control over the variables

❖ **Limitations**
  ★ Hard to find appropriate case studies
  ★ Hard to quantify findings

See:
Flyvbjerg, B.; Five Misunderstandings about Case Study Research. Qualitative Inquiry 12 (2) 219-245, April 2006
Pilot Studies

Controlled introduction of a tool/technique into a real project, where the researcher can no longer control the context, but where the net effect can be measured (e.g. against a baseline, or against previous experience)

- **Good for**
  - Measuring the benefits in a real setting
  - Preparation for tech. transfer
  - Getting organizations interested in your work

- **Limitations**
  - Hard to get organizations to adopt unproven ideas
  - Hawthorn effect (and other bias problems)

**See:**

Rational Reconstructions

A demonstration of a tool or technique on data taken from a real case study, but applied after the fact to demonstrate how the tool/technique would have worked

◇ Good for
  ★ Initial validation before expensive pilot studies
  ★ Checking the researcher's intuitions about what the tool/technique can do

◇ Limitations
  ★ potential bias (you knew the findings before you started)
  ★ easy to ignore “signal-to-noise ratio”

◇ Examples
  ★ LAS; BART; ... etc.

See:
Examples in Cohen Empirical Methods for Artificial Intelligence
Exemplars

Self-contained, informal descriptions of a problem in some application domain; exemplars are to be considered immutable; the specifier must do the best she can to produce a specification from the problem statement.

✦ Good for:
  ★ Setting research goals,
  ★ Understanding differences between research programs

✦ Limitations:
  ★ No clear criteria for comparing approaches
  ★ Not clear that “immutability” is respected in practice

✦ Examples:
  ★ Meeting Scheduler; Library System; Elevator Control System; Telephones;...

see:
  M. S. Feather, S. Fickas, A. Finkelstein, and A. van Lamsweerde,
Surveys

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

- **Good for**
  - Investigating the nature of a large population
  - Testing theories where there is little control over the variables

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

See:

Artifact / Archive Analysis

Investigation of the artifacts (documentation, communication logs, etc) of a software development project after the fact, to identify patterns in the behaviour of the development team.

- **Good for**
  - Understanding what really happens in software projects
  - Identifying problems for further research

- **Limitations**
  - Hard to build generalizations (results may be project specific)
  - Incomplete data

**See:**
Ethnographies

Interpretive, in-depth studies in which the researcher immerses herself in a social group under study to understand phenomena though the meanings that people assign to them

❖ **Good for:**
  ★ Understanding the intertwining of context and meaning
  ★ Explaining cultures and practices around tool use

❖ **Limitations:**
  ★ No generalization, as context is critical
  ★ Little support for theory building

**See:**
Action Research

Research and practice intertwine and shape one another. The researcher mixes research and intervention and involves organizational members as participants in and shapers of the research objectives

◆ Good for
  ★ Any domain where you cannot isolate variables, cause from effect, ...
  ★ Ensuring research goals are relevant
  ★ When effecting a change is as important as discovering new knowledge

◆ Limitations
  ★ Hard to build generalizations (abstractionism vs. contextualism)
  ★ Won't satisfy the positivists!

See:
Simulations

An executable model of the software development process, developed from detailed data collected from past projects, used to test the effect of process innovations

- **Good for:**
  - Preliminary test of new approaches without risk of project failure
  - [Once the model is built] each test is relatively cheap

- **Limitations:**
  - Expensive to build and validate the simulation model
  - Model is only as good as the data used to build it
  - Hard to assess scope of applicability of the simulation

- **See:**
Benchmarks

A test or set of tests used to compare alternative tools or techniques. A benchmark comprises a motivating comparison, a task sample, and a set of performance measures.

- **Good for**
  - Making detailed comparisons between methods/tools
  - Increasing the (scientific) maturity of a research community
  - Building consensus over the valid problems and approaches to them

- **Limitations**
  - Can only be applied if the community is ready
  - Become less useful / redundant as the research paradigm evolves

See:

Questions

- Do any of these idioms capture your research?
  - Do the distinctions make sense?
  - Are there other idioms we’ve missed?

- Are we (as a community) using the right idioms?
  - Should we be using some of them more than we do?
  - Should we be using some of them less than we do?

- What standards of reporting should we demand?
  - Eg, when reviewing papers for SE conferences
  - Should we be more explicit about our research methods?

- What practical steps can we take...
  - Workshops on research validation?
  - Benchmarking initiatives?
Validating SE models

- **Logical Positivist view:**
  - “There is an objective world that can be modeled by building a consistent body of knowledge grounded in empirical observation”
  - **In SE:** “there is an objective problem that exists in the world”
    - Build a consistent model; make sufficient empirical observations to check validity
    - Use tools that test consistency and completeness of the model
    - Use reviews, prototyping, etc to demonstrate the model is “valid”

- **Popper’s modification to logical positivism:**
  - “Theories can’t be proven correct, they can only be refuted by finding exceptions”
  - **In SE:** “models must be refutable”
    - Look for evidence that the model is wrong
    - Eg, collect scenarios and check the model supports them

- **Post-Modern view:**
  - “There is no privileged viewpoint; all observation is value-laden; scientific investigation is culturally embedded”
  - Eg, Kuhnian paradigms; Toulmin’s weltanschauungen
  - **In SE:** “validation is always subjective and contextualised”
    - Use stakeholder involvement so that they ‘own’ the requirements models
    - Use ethnographic techniques to understand the weltanschauungen