

Empirical Methods for Socio-Technical Research

Instructors: James D. Herbsleb / Marcelo Cataldo

Class Meeting Time: Tuesdays 9:00a-11:50a

Location: NSH 3002

University Units: 12

Office Hours: by request

Course Description.

The past decades has witness the emergence, among others, of two important trends in software systems. First, an increasing number of software technologies have a significant social component (e.g. end-user programming, collaborative development tools, phishing detection tools, etc). Second, software systems have increased significantly in complexity and size challenging traditional development and testing approaches. Empirical research methods provide a basic set of tools for researchers interested in designing and evaluating such system. This course is designed to acquaint you with several basic types of empirical methods such as ethnography, interviews, surveys and experimental and quasi-experimental design. Ultimately, the goal of the course is to develop your skills to make appropriate methodological decisions and allow you to become a better and more critical consumer of research. You will be required to critique a number of examples of published research as well as apply one or more methods to your current research projects to address an important software engineering question.

Course Overview.

The course will have the following general sequence.

- *Introduction to Empirical Methods* – An overview of the nature of empirical research, the various methodological approaches, as well as examples of good and not so good empirical studies will be discussed.
- *Qualitative Methods* – The central characteristics of qualitative research as well as traditional methodological approaches such as ethnography and interviews will be discussed.
- *Experimental and Quasi-Experimental Design* – Various design variants for experiments and quasi-experiments will be discussed and evaluated.
- *Measurement* – Measures are quite often used as the basis for empirical claims. The issues of reliability, validity and generalizability will be defined and discussed.
- *Data Collection Methods* – A collection of approaches for data collection will be examined including survey methodologies, the use of archival data, and content analysis??.
- *Implementing a Research Design* – A critical element of success in the appropriate implementation of the study whether it is archival, in the laboratory, or in the field. This particular issue will be part of the discussions throughout the entire course.

Class textbook and additional readings.

Rosenthal, R. and Rosnow, R.L.(2008). *Essentials of Behavioral Research: Methods and Data Analysis – 3rd Ed.* McGraw-Hill. (abbreviation R&R)

There are additional books that are solid references in various topics in research methods. You are welcome to acquire them but it is not required. We will only read a few chapters or part of chapters from them which will be available in electronic form in the "Course Documents" folder. The books are:

King, G., Keohane, R.O. and Verba, S. (1997). *Designing Social Inquiry.* Princeton Press. (abbreviation KKV)

Shadish, W.R., Cook, T.D. and Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Houghton Mifflin Company. (abbreviation SCC)

Strauss, A. and Corbin, J. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory - 2nd Ed.* Sage. (abbreviation S&C)

Dillman, D.A., Smyth, J.D. and Christian, L.M. (2009). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method – 3rd Ed.* John Wiley & Sons. (abbreviation DSC)

Additional papers will be assigned.

Class grading.

The students will have to prepare a 1-page critique of the assigned research papers. There will be two small assignments where the students will have to apply a particular research method to address a research question and present the research design to class. The final project will consist on a research paper addressing a software engineering research question of interest to the student. One or more empirical methods presented in class have to be part of the research paper.

The grading for the class will be structured in the following way:

- Readings critiques 30%
- Two small assignments 30% (15% each)
- Final project 40%

Paper Critiques.

Students are expected to submit critiques of published papers we will discuss in class. There are 12 papers marked with (*) and students should select 10 of those papers. Each paper critique will have a limit of 500 words and will be due before class on the day the selected paper is discussed. We will examine how to prepare paper critiques in the first few classes of the course.

Small Assignment.

We will have two small assignments in the course. Each assignment will consist on applying a research method to address one of three research questions provided by the instructors. Alternatively, students can choose to apply the research method to their own research projects.

For each assignment, students will prepare two deliverables. A report describing the research question/s, the details of the research method used and how the design allows you to answer the research questions should be prepared. The reports are due at noon on the day prior the assignment discussion class. The second deliverable consists of a 10-minute presentation to be given in class. Each presentation will be followed by a 10 minute discussion. The reports will be made available to the class and students are expected to read them prior to class.

Final Project.

An important amount of effort will be dedicated to a semester-long project in which students go through the process of articulating a set of research questions and developing the research methods appropriate for answering those questions. The outcome of the project will be a research paper that will contain 3 sections: introduction, literature review + research questions, and research method. The content is expected to be of sufficient detail and quality to be reviewable as a research proposal by a panel of experts from a funding agency. Throughout the semester, we will have deadlines for intermediate deliverable in order to motivate a continuous progress of the project.

Class Syllabus.

Class #1 - January 12th:

Topics: Introduction to Course / Motivation for Empirical Methods / Testing theory versus testing an outcome / A few examples of bad and good use of empirical methods

Readings:

- R&R - Chapter 2 (*read pages 37 through 54*)
- KKV - Chapter 1 (*read pages 1 through 28*)
- McGrath (1994). Methodology Matters: Doing Research in the Behavioral and Social Sciences. In *Readings in Human-Computer Interaction: Toward the Year 2000*. Baecker, Gruding, Buxton & Greenberg, Editors. (read pages 155-164)
- Examples of paper critiques.

Class #2 - January 19th:

Topics: The nature of qualitative research / Introduction to Ethnography and Interviews

Readings:

- R&R - Chapter 6 (*read pages 160-177*)
- S&C - Chapters 1, 2, 3 and 4.
- Cassell, C. and Symon, G. (2004) - Chapters 2 and 3.
- Dourish, P. (2006). Implications for Design. In *Proceedings of CSCW'06*, Banf, Canada.
- Cramton, C.D. (2001). The Mutual Knowledge Problem and its Consequences for Dispersed Collaboration. *Organization Science*, Vol. 12, No. 3, pp. 346-371.

Class #3 - January 26th:

Topics: Introduction to Survey Research / Questionnaire Construction / Computer Mediated Data Collection / Longitudinal Surveys

Readings:

- Warwick, D.P. and Lininger, C.A. *The Sample Survey: Theory and Practice*. Chapters 2, 3 and 6.
- DSC - Chapters 4 and 5.
- R&R - Chapter 6 (*read pages 177-186*)
- Lynn, P. (2009). *Methodology of Longitudinal Surveys*. Chapter 1.
- Hinds, P. and McGrath, C. (2006). Structures that Work: Social Structure, Work Structure and Coordination Ease in Geographically Distributed Teams. In *Proceedings of CSCW'06*, Banff, Canada.
- Kossinets, G. (2006). Effects of Missing Data in Social Networks. *Social Networks*, Vol. 28, pp. 247-268.

Class #4 - February 2nd:

Topics: Introduction to Measurement / Reliability / Validity / Generalizability / Scaling

Readings:

- Ghiselli, E.E., Campbell, J.P. and Zedeck, S. (1981). *Measurement Theory in the Behavioral Sciences*. Freeman & Company. Chapters 2 and 3.
- R&R - Chapter 4
- SCC - Chapters 2 and 3
- Cataldo et al (2006). Identification of Coordination Requirements: Implications for the Design of Collaboration and Awareness Tools. In *Proceedings of CSCW'06*, Banff, Canada.

February 9th: NO CLASS

Class #5 - February 16th:

Assignment #1 DUE on February 15th at NOON

Topics: Assignment #1 discussion

Readings: Assignment #1 submissions as basis for the class discussion

Class #6 - February 23rd:

Final Project Proposal DUE (max. 1000 words)

Topics: Introduction to Experimental Design

Readings:

- R&R - Chapters 7 and 8
- Aronson, E., Wilson, T.D. and Brewer, M. (1998). Experimentation in Social Psychology. In *Handbook of Social Psychology*, pp. 99-142.
- Dabbish, L. and Kraut, R. (2008). Awareness Displays and Social Motivation to Coordinating Communication. *Information Systems Research*, Vol. 19, No. 2, pp. 221-238. (*)
- Straus, S.G. and McGrath, J.E. (1994). Does the Medium Matter? The Interaction of Task Type and Technology on Group Performance and Member Reaction. *Journal of Applied Psychology*, Vol. 79, No. 1, pp. 87-97. (*)

Class #7 – March 2nd:

Topics: The Role of Statistics / Understanding Interactions

Readings:

- R&R - Chapter 12
- Abelson, R.P. (1995). *Statistics as a Principled Argument*, LEA. Chapter 1.
- Baron, R.M. & Kenny, D. (1986). The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical considerations. *Journal of Personality and Social Psychology*, Vol. 51, No. 6, pp. 1173-1182.

March 9th: NO CLASS – SPRING BREAK

Class #8 - March 16th:

Topics: Quasi-Experimental Design

Readings:

- SCC - Chapter 4
- SCC - Chapter 5 (*read pages 135-156*)
- SCC - Chapter 6 (*read pages 175-181*)
- Teasley, S.D. et al (2002). Rapid Software Development through Team Collocation. *IEEE Transactions in Software Engineering*, Vol. 28, No. 7, pp. 671-683. (*)

Class #9 - March 23rd:

Final Project “Intro + Literature Review + brief outline of research method” DUE

Topics: Introduction to Doing Lab Experiments / Problems and challenges

Readings:

- R&R - Chapter 9
- SCC - Chapters 9 and 10
- Mook, D. (1983). In Defense of External Invalidity. *American Psychologist*, Vol. 38, No. 4, pp. 379-387.
- Sears, D. (1986). College Sophomores in the Laboratory: Influences of a Narrow Database on Social Psychology's View of Human Nature. *Journal of Personality and Social Psychology*, Vol. 51, No. 3, pp. 515-530.

Class #10 - March 30th:

Assignment #2 DUE on March 29th at NOON

Topics: Assignment #2 discussion

Readings: Assignment #2 submissions as basis for the class discussion

Class #11 - April 6th:

Topics: Writing Scientific Papers

Readings:

- Bem, D.J. (2003). Writing the Empirical Journal Article. In *The Complete Academic: A Practical Guide for the Beginning Social Scientist*, Darley, Zanna and Roediger, Editors.
- Lanham, R. McSweeney's Internet Tendency: Internet-Age Writing. [<http://www.mcsweeneys.net/2009/4/20lanham.html>]
- Mahoney, M. (1985). Open exchange and epistemic progress. *American Psychologist*, Vol. 40, No. 1, pp. 29-39.
- Stinchcombe, A.L. & Ofshe, R. (1986). *On Journal Editing as a Probabilistic Process*, Cambridge University Press. pp 333-336.
- Acceptance and Rejection:
 - Case I: Cataldo et al from ICSE/CSCW
 - Case II: Frost & Taylor from Administrative Science Quarterly

Class #12 - April 13th:

Topics: Exploratory Data Analysis

Readings:

- R&R - Chapter 10
- Wainer, H. (1984). How to Display Data Badly. *The American Statistician*, Vol. 38, No. 2, pp. 137-147.
- Wainer, H. and Vellman, P.F (2001). Statistical Graphics: Mapping the Pathways of Science. *Annual Review of Psychology*, Vol. 52, pp. 305-335.
- Moody, J. et al (2005). Dynamic Network Visualizations. *American Journal of Sociology*, Vol. 110, No. 4, pp. 1206-41. (*)
- Shen, Z. et al (2006). Visual Analysis of Large Heterogeneous Social Networks by Semantic and Structural Abstraction. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1-13. (*)

Class #13 - April 20th:

Topics: Analysis of Large-Scale Datasets / Analysis of Trace Data

Readings:

- Kamvar, S and Harris, J. An Almanac of Human Emotions, <http://wefeelfine.org>
- Salzberg, S.L. (1997). On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach. *Data Mining and Knowledge Discovery*, Vol. 1, pp. 317-328.
- Boh, W.F. et al. (2007). Learning from Experience in Software Development: A Multi-level Analysis. *Management Science*, Vol. 53, No. 8, pp. 1315-1331. (*)
- Gilbert, E. and Karahalios, K. (2009). Predicting tie strength with social media. In *Proceedings of CHI'09*, Boston, USA. (*)
- Hahn, J., Moon, J.Y., Zhang, C. (2008). Emergence of New Project Teams from Open Source Software Developer Networks: Impact of Prior Collaboration Ties. *Information Systems Research*, 19, 3, pp. 369-391. (*)

Class #14 - April 27th:

Topics: Multi-Method Research

Readings:

- Wang, H., Fussell, S.R. and Setlock, S.L. (2009). Cultural Differences and Adaptation of Communication Styles in Computer-Mediated Group Brainstorming. In *Proceedings of CHI'09*, Boston, USA. (*)
- Levitt, R.E. et al (1994). The Virtual Design Team: Simulating How Organization Structure and Information Processing Tools Affect Team Performance. In *Computational Organization Theory*, K. M. Carley & M.J. Prietula, Editors. (*)
- Siggelkow, N. (2002). Evolution Toward Fit. *Administrative Science Quarterly*, Vol. 47, pp. 125-159. (*)

- MacCormack, A. and Verganti, R. (2003). Managing the Sources of Uncertainty: Matching Process and Context in Software Development. *The Journal of Product Innovation Management*, Vol. 20, pp. 217-232. (*)

Class #15 - May 4th:

Topics: Final project presentations

No Readings

May 7th: Final Project Report DUE at NOON