The 1-Page Summary
(of a 35-page paper)

- Hypothesis generation – measurable quantities
- Stages of an experiment
  - Conception, design, preparation, execution, analysis, dissemination/decision making
- Experimental Design
  - Number of factors, interactions
  - Nested vs Crossed
- Refresh of statistics
  - T-tests, F-tests, Correlations, Parametric tests, Regressions
Points of Discussion

- Accuracy of surrogate measures
- "Controlled" factors
  - experience, environment, relationships, blocking
- Post-Positivist vs. Constructivist
  - hypotheses generation
  - formal experiments or case studies or ???
- Statistics test
  - hypotheses restricted to testing differences
  - not as straight forward as it seems

Formal Methods Application:
An Empirical Tale of Software Development
(with comments and a rebuttal)

Sobel and Clarkson (2002);
Berry and Tichy (2003);
**The Study Summary**  
*(Sobel and Clarkson 2002)*

- **Experimental unit**: teams of students (average 2 / team)
- **Control group**: “no” formal methods instruction
- **Experimental group**: formal methods instruction

**Hypotheses:**
- Formal methods instruction increases complex problem solving skills
- Formal methods instruction results in “better” programs

**Design overview:**
- Control (13 teams) / Experimental (6 teams)
- Experimental group participants self-assigned
- Metrics of “better”
  - Code correctness – executable performance in 6 test cases
  - Conciseness, complexity – qualitative analysis of artifacts

---

**The Comments Summary**  
*(Berry and Tichy 2003)*

- Claim the study was not well designed
- Suggest conclusions drawn by Sobel and Clarkson are invalid

**Critique:**
- **Internal validity**
  - Subject groups were not matching
  - Differing pre- and post-tests (based on other paper)
  - No post-treatment questionnaire
- **Analysis**
  - Test-cases not described / too few administered (stat. testing)
  - Artifacts were not mandatory
- **External validity**
  - No discussion of threats to validity

- Argument leverages details of same study from a prior paper by the same authors
The Rebuttal Summary
(Sobel and Clarkson 2003)

Argue that Berry and Tichy misunderstood the experiment
- The design was sound for a quasi-experiment
- The experiment was not designed for a laboratory

Hypotheses studied should be considered in context of the larger, three-year study

Only motivation and novelty are plausible confounds; all other confounds were reasonably controlled

Given the academic setting, more controlled measurement was simply not possible

All other identified issues were not severe enough to invalidate the study

Discussion

Were the criticisms of Berry and Tichy fair and grounded? Did Sobel and Clarkson convincingly defend their experimental design?

Was the Sobel and Clarkson study well designed?
- Participant/group selection: motivation, skill level, learning style, student maturity…
- Experimental unit was team, but conclusions drawn for individuals
- “absence of proof is not proof of absence”; is failure to collect relevant data bad experimental design?

If you were planning this study, would you change the experimental design?
Discussion

- Should reporting on an experiment that is part of a larger study be held to different standards than an independent experiment?

- How clearly should an author state rationale for experimental design, methods, and analysis? Can some details be assumed to be implied/understood?