Writing Good Software Engineering Research Papers

Mary Shaw
Carnegie Mellon University

www.cs.cmu.edu/~shaw/
Good Writing Needs Good Content

• Writing a good paper depends on having good research to write about
  > If the result is not significant, it doesn’t matter how good the paper is
  > If your claims don’t match your results, you’ll have trouble providing convincing evidence

• It’s also hard work, a skill that requires practice. Writing a paper is like designing a system.

• So this minitutorial addresses both your research strategy and how you present the work
Plan

• Life cycle of a technological innovation
  > Different issues, venues at different stages

• Focus on research papers
  > Various authors, conference advice

• Elements of a research presentation
  > Question, result, validation
  > Data from ICSE 2002, 2003

• Research strategies that work
  > The logical structure of a project and paper
  > Examples from ICSE 2003
Redwine/Riddle Maturation Model

- **Basic Research**
  - Recognize problem
  - Invent ideas

- **Concept Formation**
  - Refine ideas
  - Publish solutions

- **Development & Extension**
  - Try it out
  - Clarify
  - Refine

- **Internal Exploration**
  - Stabilize
  - Port
  - Use for real problems

- **External Exploration**
  - Broaden user group
  - Extend

- **Popularization**
  - Propagate through community

---

Seminal paper or system

Usable capability

Outsiders use it

Production quality, commercial support

Key Idea

Sam Redwine, Jr. and William Riddle: *Software Technology Maturation*, Proc ICSE-8, May 1985
<table>
<thead>
<tr>
<th>Major Technology Areas</th>
<th>58</th>
<th>59</th>
<th>60</th>
<th>61</th>
<th>62</th>
<th>63</th>
<th>64</th>
<th>65</th>
<th>66</th>
<th>67</th>
<th>68</th>
<th>69</th>
<th>70</th>
<th>71</th>
<th>72</th>
<th>73</th>
<th>74</th>
<th>75</th>
<th>76</th>
<th>77</th>
<th>78</th>
<th>79</th>
<th>80</th>
<th>81</th>
<th>82</th>
<th>83</th>
<th>84</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software Engineering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compiler Construction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge-Based Systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metrics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology Concepts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structured Programming</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abstract Data Types</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methodology Technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCR Methodology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOD-STD-SDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFR 800-14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consolidated Technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smalltalk 80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost Models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unix</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SREM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Software Technology Maturation Points

<table>
<thead>
<tr>
<th>Basic Resch</th>
<th>Concept Form</th>
<th>Dev + Ext</th>
<th>Internal Exp</th>
<th>External Expl</th>
<th>Popularize</th>
</tr>
</thead>
</table>

Maturation Times

<table>
<thead>
<tr>
<th>Module</th>
<th>Maturation Times</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFR 800-14</td>
<td></td>
</tr>
<tr>
<td>unix</td>
<td></td>
</tr>
<tr>
<td>Structured Programming</td>
<td></td>
</tr>
<tr>
<td>Verification</td>
<td></td>
</tr>
<tr>
<td>Abstract Data Types</td>
<td></td>
</tr>
<tr>
<td>SREM</td>
<td></td>
</tr>
<tr>
<td>Compiler Construction</td>
<td></td>
</tr>
<tr>
<td>SCR Methodology</td>
<td></td>
</tr>
<tr>
<td>Smalltalk 80</td>
<td></td>
</tr>
<tr>
<td>Cost Models</td>
<td></td>
</tr>
<tr>
<td>Software Engineering</td>
<td></td>
</tr>
<tr>
<td>Metrics</td>
<td></td>
</tr>
<tr>
<td>Knowledge-Based Systems</td>
<td></td>
</tr>
<tr>
<td>DOD-STD-STD</td>
<td></td>
</tr>
</tbody>
</table>

- Basic Resch: 6 years
- Concept Form: 4 years
- Dev + Ext: 4 years
- Internal Exp: 3 years
- External Expl: Popula
Phase Times and Publications

<table>
<thead>
<tr>
<th>Basic Resch</th>
<th>Concept Form</th>
<th>Dev + Ext</th>
<th>Internal Exp</th>
<th>External Pop</th>
<th>Popularize</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
</tr>
</tbody>
</table>

Typical publication venues

- Research workshops
- Conferences
- Archival journals
- Reviews
- Development workshops
- Popular journals
- Trade publications
Success needs cumulative evidence

• A single paper has limited scope
  > Conference papers can hold one idea
  > Journal papers can wrap up individual results

• Results are more convincing if they are confirmed in different ways (triangulation)

• Each promising step justifies investment in next (often more expensive) step
Plan

• Life cycle of a technological innovation
  > Different issues, venues at different stages

• Focus on research papers
  > Various authors, conference advice

• Elements of a research presentation
  > Question, result, validation
  > Data from ICSE 2002, 2003

• Research strategies that work
  > The logical structure of a project and paper
  > Examples from ICSE 2003
Research Styles

• Physics and medicine have well-recognized research styles
  > Hypothesis, controlled experiment, analysis, refutation
  > Double-blind large-scale studies
• Acceptance of results relies on process as well as analysis
• Simplified versions help to explain the field to observers
  »
• Fields can be characterized by identifying what they value:
  > What kinds of questions are “interesting”?  
  > What kinds of results help to answer these questions?  
  » What research methods can produce these results?  
  > What kind of evidence demonstrates the validity of a result?
Critiques of Experimental CS/SE

“Computer scientists publish relatively few papers with experimentally validated results … The low ratio of validated results appears to be a serious weakness in CS research. This weakness should be rectified”

• Studies over past few years criticize computer science for failure to collect, report, analyze experimental data

• They start with the premise that data must be collected, then analyze papers and find data lacking

• I ask a different question: What are the characteristics of software engineering research that the field recognizes as quality research?

Newman: Pro Forma Abstracts

- Asked, “To what extent is HCI an engineering discipline”?
- Characterized engineering research products
- Created three pro forma abstracts, templates describing research
- 90% of papers in engineering research fit these templates

EM: Enhanced model

Existing model-type models are deficient in dealing with properties of solution strategy. An enhanced model-type is described, capable of providing more accurate analyses / predictions of properties in solution strategy designs. The model has been tested by comparing analyses / predictions with empirically measured values of properties.

ES: Enhanced solution

Studies of existing artifact-type have shown deficiencies on property. An enhanced design for an artifact-type is described, based on solution strategy. In comparison with existing solutions, it offers enhanced levels of property, according to analyses based on model-type. These improvements have been confirmed / demonstrated in tests of a working artifact-type based on the design.

ET: Enhanced tool

The effectiveness of model-type / solution strategy in supporting the design of artifact-type has been demonstrated. An enhanced tool / method is described for the design of artifact-type based on model-type / solution strategy. Examples are provided confirming the effectiveness of its support for model-type / solution strategy in design.
Newman: Pro Forma Abstracts

- Only 25-30% of HCI papers fit
- Created 2 more pro forma abstracts (arguably engineering)
- Now 95% of HCI papers fit
- Notes
  - Preliminary study, e.g., no check on inter-rater reliability
  - Found this a useful device for reading papers
  - Influenced refereeing in CHI

RS: Radical solution

A radical solution to the problem of problem definition is described, based on solution strategy. In comparison with existing normal solutions it offers advantages, which have been demonstrated in preliminary tests, but it leaves a number of side effects to be addressed including list of side effects. Strategies are suggested for addressing these side effects.

XH: Experience and/or Heuristic

Studies reported here of application supported by supporting technology generate a number of findings concerning issues, including list-of-findings. They indicate that requirement is / is not met by design-heuristic.
Brooks proposed recognizing three kinds of results, with individual criteria for quality:

> **findings** -- well-established scientific truths -- judged by truthfulness and rigor

> **observations** -- reports on actual phenomena -- judged by interestingness

> **rules-of-thumb** -- generalizations, signed by an author (but perhaps not fully supported by data) -- judged by usefulness

with freshness as criterion for all
Conference-specific advice

• There’s lots of “how to write a paper” advice
  > OOPSLA, POPL, PLDI, SOSP, SIGCOMM, SIGGRAPH
  > Links on my writing advice web site
    » www.cs.cmu.edu/~shaw > Education > WordWright
    » Under Resources > CS Advice

• HCI community does better
  > Newman analysis above
  > Analysis of regional differences in acceptance rates
Plan

- Life cycle of a technological innovation
  > Different issues, venues at different stages
- Focus on research papers
  > Various authors, conference advice
- Elements of a research presentation
  > Question, result, validation
  > Data from ICSE 2002, 2003
- Research strategies that work
  > The logical structure of a project and paper
  > Examples from ICSE 2003
Key objectives

- **Quality** -- utility as well as functional correctness
- **Cost** -- both of development and of use
- **Timeliness** -- good-enough result, when it’s needed

• Address problems that affect practical software
## Types of Research Questions

<table>
<thead>
<tr>
<th>Category</th>
<th>Example Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method/means of development</td>
<td>How can we do/create/automate X? What is a better way to do/create X?</td>
</tr>
<tr>
<td>Method for analysis</td>
<td>How can I evaluate the quality of X? How do I choose between X and Y?</td>
</tr>
<tr>
<td>Evaluation / analysis of an instance</td>
<td>What is property X of artifact/method Y? How does X compare to Y? What is the current state of X / practice of Y?</td>
</tr>
<tr>
<td>Generalization / characterization</td>
<td>Is X always true of Y? Given X, what is Y? What, exactly, do we mean by X? Is Y a good formal/empirical model for X? What are the types of X, how are they related?</td>
</tr>
<tr>
<td>Feasibility</td>
<td>Does X exist, and what is it? Is it possible to do X at all?</td>
</tr>
</tbody>
</table>
### ICSE 2002 submissions

**Graphs:**
- **Question Distribution:**
  - **Devel:** Accepted: 50, Rejected: 0, Total: 50
  - **Analy:** Accepted: 100, Rejected: 0, Total: 100
  - **Eval:** Accepted: 150, Rejected: 0, Total: 150
  - **Gener:** Accepted: 200, Rejected: 0, Total: 200
  - **Feas:** Accepted: 250, Rejected: 0, Total: 250
  - **Total:** Accepted: 900, Rejected: 0, Total: 900

- **Percentage Distribution:**
  - **Devel:** Accepted: 0%, Rejected: 0%, Total: 0%
  - **Analy:** Accepted: 20%, Rejected: 0%, Total: 20%
  - **Eval:** Accepted: 40%, Rejected: 0%, Total: 40%
  - **Gener:** Accepted: 60%, Rejected: 0%, Total: 60%
  - **Feas:** Accepted: 80%, Rejected: 0%, Total: 80%
  - **Total:** Accepted: 100%, Rejected: 0%, Total: 100%

<table>
<thead>
<tr>
<th>Type of question</th>
<th>Submitted</th>
<th>Accepted</th>
<th>2003</th>
<th>Ratio Acc/Sub</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method or means of development</td>
<td>142(48%)</td>
<td>18(42%)</td>
<td>13</td>
<td>(13%)</td>
</tr>
<tr>
<td>Method for analysis or evaluation</td>
<td>95(32%)</td>
<td>19(44%)</td>
<td>18</td>
<td>(20%)</td>
</tr>
<tr>
<td>Design, evaluation, or analysis of a particular instance</td>
<td>43(14%)</td>
<td>5(12%)</td>
<td>4</td>
<td>(12%)</td>
</tr>
<tr>
<td>Generalization or characterization</td>
<td>18(6%)</td>
<td>1(2%)</td>
<td>7</td>
<td>(6%)</td>
</tr>
<tr>
<td>Feasibility study or exploration</td>
<td>0(0%)</td>
<td>0(0%)</td>
<td>0</td>
<td>(0%)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>298(100.0%)</td>
<td>43(100.0%)</td>
<td>42</td>
<td>(14%)</td>
</tr>
</tbody>
</table>
What do PCs look for?

• Clear statement of the question you answered
  > that is, the problem about software you answered
• Explanation of why the problem matters
<table>
<thead>
<tr>
<th>Types of Research Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure / technique</td>
</tr>
<tr>
<td>Qualitative or descr. model</td>
</tr>
<tr>
<td>Analytic model</td>
</tr>
<tr>
<td>Empirical model</td>
</tr>
<tr>
<td>Tool / notation</td>
</tr>
<tr>
<td>Specific solution</td>
</tr>
<tr>
<td>Report</td>
</tr>
</tbody>
</table>
ICSE 2002 submissions

<table>
<thead>
<tr>
<th>Type of result</th>
<th>Submitted</th>
<th>Accepted</th>
<th>2003 Ratio Acc/Sub</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure or technique</td>
<td>152(44%)</td>
<td>28(51%)</td>
<td>18</td>
</tr>
<tr>
<td>Qualitative or descriptive model</td>
<td>50(14%)</td>
<td>4 (7%)</td>
<td>7</td>
</tr>
<tr>
<td>Empirical model</td>
<td>4 (1%)</td>
<td>1 (2%)</td>
<td>5</td>
</tr>
<tr>
<td>Analytic model</td>
<td>48(14%)</td>
<td>7 (13%)</td>
<td>11</td>
</tr>
<tr>
<td>Tool or notation</td>
<td>49(14%)</td>
<td>10(18%)</td>
<td>5</td>
</tr>
<tr>
<td>Specific solution, prototype, answer, or judgment</td>
<td>34(10%)</td>
<td>5 (9%)</td>
<td>2</td>
</tr>
<tr>
<td>Report</td>
<td>11(3%)</td>
<td>0 (0%)</td>
<td>1</td>
</tr>
<tr>
<td>TOTAL</td>
<td>348(100.0%)</td>
<td>55(100.0%)</td>
<td>49</td>
</tr>
</tbody>
</table>
What do PCs look for?

• What’s new? How is it related to prior work?
• What, precisely, does the research claim to show?
  > If it should work on large systems, show it scales
  > If it’s “automatic”, don’t use manual intervention
  > If it’s “distributed”, don’t assume central server
  > If it’s a new notation, show why it’s better
  > If it’s a new model, be clear about its power
  > If it’s a new design element, treat it as a generalization
  > If it’s a synthesis, say why the synthesis is novel
  > If an implementation is featured, show its role
# Types of Research Validation

## Analysis
- Formal model: rigorous derivation and proof
- Empirical model: data on use in controlled situation
- Controlled experiment: carefully designed statistical experiment

## Experience
- Qualitative model: narrative
- Empirical model, tool: data, usually statistical, on practice
- Notation, technique: comparison of systems in actual use

## Example
- Here’s how my result works on a small example

## Evaluation
- Descriptive model: adequately describes phenomena of interest
- Empirical model: is able to predict … because …

## Persuasion
- I thought hard about this, and I believe...

## Blatant assertion
- No serious attempt to evaluate result
ICSE 2002 submissions

<table>
<thead>
<tr>
<th>Type of validation</th>
<th>Submitted</th>
<th>Accepted</th>
<th>2003 Ratio Acc/Sub</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>48(16%)</td>
<td>11(26%)</td>
<td>11</td>
</tr>
<tr>
<td>Evaluation</td>
<td>21(7%)</td>
<td>1 (2%)</td>
<td>7</td>
</tr>
<tr>
<td>Experience</td>
<td>34(11%)</td>
<td>8 (19%)</td>
<td>7</td>
</tr>
<tr>
<td>Example</td>
<td>82(27%)</td>
<td>16(37%)</td>
<td>17</td>
</tr>
<tr>
<td>Some example, can't tell whether it's toy or actual use</td>
<td>6 (2%)</td>
<td>1 (2%)</td>
<td>0</td>
</tr>
<tr>
<td>Persuasion</td>
<td>25(8%)</td>
<td>0 (0.0%)</td>
<td>0</td>
</tr>
<tr>
<td>No mention of validation in abstract</td>
<td>84(28%)</td>
<td>6 (14%)</td>
<td>-</td>
</tr>
<tr>
<td>TOTAL</td>
<td>300(100.0%)</td>
<td>43(100.0%)</td>
<td>42</td>
</tr>
</tbody>
</table>
What do PCs look for?

• Solid evidence: why the reader should believe result
• Validation related to the claim
  > If you improve on prior art, do comparison
  > If you did analysis, follow its rules
  > If you cite practical experience, separate your effect
• Accurate description of the evidence
  > “case study” & “experiment” >> data & anecdotes
Commonest Types of ICSE 2002 Papers

• Question
  > Most common: improved method or means of developing software
  > Also fairly common: papers about methods for analysis, principally analysis of correctness (most common in 2003)

• Result
  > Most common: a new procedure or technique for some aspect of software development
  > Not unusual: a new analytic model

• Validation
  > Most common: analysis and experience in practice
  > Also fairly common: example idealized from practice
  > Common in submissions but not acceptances: persuasion
<table>
<thead>
<tr>
<th>Question</th>
<th>Strategy/Result</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devlpmt method</td>
<td>Proc/technique</td>
<td>Analysis</td>
</tr>
<tr>
<td>Analysis method</td>
<td>Qual/desc model</td>
<td>Experience</td>
</tr>
<tr>
<td>Evaluate instance</td>
<td>Analytic model</td>
<td>Example</td>
</tr>
<tr>
<td>Generalization</td>
<td>Empirical model</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Feasibility</td>
<td>Tool/notation</td>
<td>Persuasion</td>
</tr>
<tr>
<td></td>
<td>Specific solution</td>
<td></td>
</tr>
</tbody>
</table>
Plan

- Life cycle of a technological innovation
  - Different issues, venues at different stages
- Focus on research papers
  - Various authors, conference advice
- Elements of a research presentation
  - Question, result, validation
  - Data from ICSE 2002, 2003
- Research strategies that work
  - The logical structure of a project and paper
  - Examples from ICSE 2003
Complete Research Result

Real World
Practical problem

Research Setting
Idealized problem

Research product
(technique, method, model, system, …)

Validation Task 1:
Does the product solve the idealized problem?

Validation Task 2:
Does the result help to solve the practical problem?

Real World
Solution to practical problem

Real World
Solution to idealized problem
Two Common Plans

### Question
- Can X be better?
- Can X tell you Y?
- Evaluate instance
- Generalization
- Feasibility

### Strategy/Result
- New method
- Qual/desc model
- Analytic model
- Empirical model
- Tool/notation
- Specific solution
- Report

### Validation
- Analysis
- Report actual use
- Careful examples
- Evaluation
- Persuasion

**Question (Analysis method):** How can we automatically verify that a finite state machine specification is a safe abstraction of a C procedure?

**Result (Technique, supported by tool):**
- Extract finite model from C source code (using predicate abstraction and theorem proving); show conformance via weak simulation.
- Decompose verification to match software design so results compose.
- Tool interfaces with public theorem provers

**Validation (Examples):**
- Use examples whose correct outcome is known
- Compare performance with various public provers incorporated
- Verify OpenSSL handshake
Two Common Plans

**Question**
- Can X be better?
- Can X tell you Y?
- Evaluate instance
- Generalization
- Feasibility

**Strategy/Result**
- New method
- Qual/desc model
- Analytic model
- Empirical model
- Tool/notation
- Specific solution
- Report

**Validation**
- Analysis
- Report actual use
- Careful examples
- Evaluation
- Persuasion

**Question** (Development method): How can we improve on the traditional approach to document authoring?

**Result** (Technique):
- Document authored by team in series of workshops
- Workshops are highly structured around concrete issues

**Validation** (Experience):
- In use in Nokia since 2000
- Self-assessment by survey in 2001, good results
  - reduces calendar time for document
  - improves communication
  - reduces defects
Empirical Validation

**Question**

- Development method
- Can we predict cost?
- Evaluate instance
- Generalization
- Feasibility

**Strategy/Result**

- Cost estimation method
- Qualitative/descriptive model
- Analytic model
- Empirical model
- Tool/notation
- Specific solution
- Report

**Validation**

- Experience
- Example
- Evaluation
- Persuasion

*Statistical comparison*

**Question (Analysis method):** Can we estimate costs of developing web applications?

**Result (Technique):**
- Tailor existing COBRA method for web applications
- Get data set from web development company

**Validation (Analysis, statistically valid):**
- Establish evaluation criteria through interviews
- Apply tailored COBRA, least squares, and company’s informal model
- Compare results in several ways, including t-tests
<table>
<thead>
<tr>
<th>Question</th>
<th>Strategy/Result</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devlpmt method</td>
<td>Proc/technique</td>
<td>Analysis</td>
</tr>
<tr>
<td>Analysis method</td>
<td>Analytic model</td>
<td>Example</td>
</tr>
<tr>
<td>Evaluate instance</td>
<td>Empirical model</td>
<td>Evaluation</td>
</tr>
<tr>
<td>What do we mean by X?</td>
<td>Tool/notation</td>
<td>Persuasion</td>
</tr>
<tr>
<td>Feasibility</td>
<td>Specific solution</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Report</td>
<td></td>
</tr>
</tbody>
</table>

*Careful generalization*

*Report actual use*

**Question** (Generalization): What are benchmarks, in general, and how could using them improve software engineering research?

**Result** (Qualitative model):
- Examine three successful benchmarks
- Formulate descriptive theory
- Describe how theory should inform practice

**Validation** (Experience):
- Apply theory to interpret two reverse engineering benchmarks
- Identify three areas that are ripe for benchmarking
# A Common, but Bad Plan

An **Uncommon, but Good, Plan**

<table>
<thead>
<tr>
<th>Question</th>
<th>Strategy/Result</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Can X be better?</strong></td>
<td><strong>New method</strong></td>
<td>Analysis</td>
</tr>
<tr>
<td>Analysis method</td>
<td>Qual/desc model</td>
<td>Experience</td>
</tr>
<tr>
<td>Evaluate instance</td>
<td>Empirical model</td>
<td>Example</td>
</tr>
<tr>
<td>Generalization</td>
<td>Tool/notation</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Feasibility</td>
<td>Specific solution</td>
<td><strong>Look, it works!</strong></td>
</tr>
<tr>
<td></td>
<td>Report</td>
<td></td>
</tr>
</tbody>
</table>
Sometimes a breakthrough (but sometimes nonsense)

<table>
<thead>
<tr>
<th>Question</th>
<th>Strategy/Result</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devlpmt method</td>
<td>Proc/technique</td>
<td>Analysis</td>
</tr>
<tr>
<td>Analysis method</td>
<td>New approach</td>
<td>Experience</td>
</tr>
<tr>
<td>Evaluate instance</td>
<td>Analytic model</td>
<td>Example</td>
</tr>
<tr>
<td>New assumptions</td>
<td>Empirical model</td>
<td>Evaluation</td>
</tr>
<tr>
<td>Feasibility</td>
<td>Tool/notation</td>
<td>Persuasion</td>
</tr>
<tr>
<td></td>
<td>Specific solution</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Report</td>
<td></td>
</tr>
</tbody>
</table>
# ICSE 2002 and 03 Paper Types

<table>
<thead>
<tr>
<th></th>
<th>Devel Meth</th>
<th>Anal-Meth</th>
<th>Inst-ance</th>
<th>Gener Feas-aliz’n ability</th>
<th>Anal-ysis</th>
<th>Exper-iene</th>
<th>Exam-ple</th>
<th>Evalua-tion</th>
<th>Persua-soion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proc, Tech</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qual Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emp Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anal Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nota-tion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spec Soln</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Report</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
EM: Enhanced model

Existing model-type models are deficient in dealing with properties of solution strategy. An enhanced model-type is described, capable of providing more accurate analyses / predictions of properties in solution strategy designs. The model has been tested by comparing analyses / predictions with empirically measured values of properties.

Key: EM provides new or better way of looking at problems

Question

Generalization / characterization: What, exactly do we mean by X? What is a good formal/empirical model of X?

Result

Models, preferably analytic or empirical, but precise descriptive or qualitative are acceptable

Validation

Empirical analysis, controlled experiment; perhaps experience
Locating the *pro forma* abstracts in research strategy space

<table>
<thead>
<tr>
<th>Devel Meth</th>
<th>Anal Meth</th>
<th>Instance</th>
<th>Gener- aliz’n</th>
<th>Feas- ibility</th>
<th>Anal- ysis</th>
<th>Exper- ience</th>
<th>Exam- ple</th>
<th>Eval- uation</th>
<th>Persu- asion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proc, Tech Qual Model</td>
<td>ET</td>
<td>ET</td>
<td>EM</td>
<td>EM</td>
<td>EM</td>
<td>EM</td>
<td>EM</td>
<td>EM</td>
<td>ET</td>
</tr>
<tr>
<td>Proc, Tech Qual Model</td>
<td>ET</td>
<td>ET</td>
<td>EM</td>
<td>EM</td>
<td>EM</td>
<td>EM</td>
<td>EM</td>
<td>EM</td>
<td>ET</td>
</tr>
</tbody>
</table>

*Pro Forma* Research Strategies
Putting the Words on Paper

• A research paper is a purposeful, designed artifact
  > Just like a software system

• Apply software design techniques to paper design
  > Start with the requirement: read the call for papers
  > Select an architecture: plan the sections, what they say
  > Plan a schedule: allow time for review, revision
  > Check consistency: type-check text like code

• See writing guidance at
  > [www.cs.cmu.edu/~shaw](http://www.cs.cmu.edu/~shaw) > Education > WordWright
Examine the kinds of research questions software engineers ask and the ways they study those questions

• Research questions are of different kinds
  Kinds of interesting questions change as ideas mature

• Research strategies also vary
  They should be selected to match the research questions

• Ideas mature over time
  They grow from qualitative and empirical understanding to precise and quantitative models

• Good papers are steps toward good results
  Each paper provides some evidence, but overall validation arises from accumulated evidence
Final word – about this report

- In Brooks’ sense, a **rule of thumb** or **generalization**
- Not a technical result (a **finding**) …
  - No attempt to show anyone else can apply the model
  - No principled analysis
  - Limited data
    - one full set of abstracts and observation of PC
    - one set accepted papers as published
  - Use of abstracts as proxies for full papers is suspect
    - Though accepted 2003 papers suggest they’re not bad
  - Little discussion of related work