Contents I.

- Context of this review
  - Reasons for this review
  - Resources and their exploration
- Research techniques and principles
  - Research “methodology”
  - Research context
  - Research design
  - Data collecting and analysis
  - Results reporting, interpretation and drawing conclusions
  - Faceted paper classification scheme
  - Recommended resources
- References
- Excerpts from literature
Contents II.

- Context of this review
  - Reasons for this review
  - Resources and their exploration
- Context of SE research
  - Goals and history of SE research
  - Top research organizations, lead experts and top journals
  - Problems in SE research
  - Suggested solutions to the problems
  - Conclusions
- References
The Context of this Review

Why and how have we done this?
Reasons and goals of this review

- Creation of the first **resource and knowledge basis** for SE empirical research
- **Broad exploration** to find top organizations, experts and guidelines for SE research
- Before conducting our own studies **deeper research** into the particular method and its specifics is **needed**
Resources and their exploration

- **Google, ACM Digital Library, Springer Link, IEEE Digital Library**
- Finding organisations -> primary list of publications with respect to their relevance by abstracts
- Sorting publications by their average **Google Scholar and Web of Science citations count per year**
- Reading -> checking **references** -> adding new publications
- Repeated checking of the **organizations’ websites** for new publications
- Repeating the last three steps
- If it seems right, adding another organization
- Ending when the **top six** relevant and available publications were read
- Existing and relevant publications > found ones (cca 100 - 120) > found to be useful (cca 40-80) > available (cca 50-60) > read (11)
Research Techniques and Principals

How to do SE research?
Research “methodology”

1. Research context
2. Research design
3. Study execution
4. Data collection
5. Data analysis
6. Results presentation
7. Results interpretation, drawing conclusions

- The point of every study is (or should be) the exploration or description of the relations between representatives of four “archetype classes” – Actor, Technology, Activity, Software System [10]
Research context

- Historical background
- Researched phenomena
- Industrial context (entities, attributes and measures)
- Origin and manifestation of the issue
- Terminology
- Prior related studies (and how they are related)
- Questions previously unanswered
- Hypothesis (theory from which it is derived, which questions are addressed and how)
Hypothesis

• **One or more questions** we want to adress
  Integral part of a theory – input to **theory building**
• **Insightful** questions [7]
  ▪ Example: “Do independently-developed modules fail independently?” (in N-Version programming)
    ▪ They do not -> N-Version programming doesn't deliver on its promise of high reliability
  ▪ Counterexample: “Is object-oriented programming effective?”
    ▪ Important but not insightful! (small number of studies can't be expected to answer it)
• **Two levels of hypothesis statements** [7]
  ▪ Abstract, high level, natural language statements in everyday terms
    ▪ “meetings are an indispensable part of the inspection process”
  ▪ Concrete hypothesis in terms of the study's design
    ▪ “teams who do inspections with meetings find more defects than teams who do inspections without them”
As simple (or fully analyzed) as possible – otherwise include expert

Description of techniques (approach and methods) used

Description of studied variables and threats to validity

Randomization

Population of subjects and objects

Sampling (process, precalculation of sample size)

Definition of treatments

Assigning treatments (process)
Experimental unit (vs. unit of observation)
- Level of blinding
- Vested interest, minimize bias
- Methods of data collection (outcome measures)
- Analytical methods used
- Justification of choices made (all of them!)
- Study context (vs. industry context)
What do you want to do with the researched phenomena / technology?

1. Descriptive
2. Evaluative
3. Formulative (create it)

Subsets of 1
- Explanatory, Informational, Analytical

Subsets of 2
- Normative (judge it)

Subsets of 3
- Developmental, Propositional

Others
- Deductive, Explorative
Research methods - primary

- How do you want to do it (through what means)?
- **Experiment (true/controlled)**
- **Case study**
- **Survey**
- **Action research**
- **Field study/experiment**
- **Observational study/experiment**
- **Quasi-experiment (experiment not randomized)**
- **Post-mortem analysis, conceptual analysis, concept implementation (proof of concept), data analysis, instrument development, mathematical proof, protocol analysis, simulation**
Research methods - secondary

- Replication
- Meta-analysis
  - Generalization of related but not similar studies
- Systematic review
- Literature review
  - Review only – should not be normative
- Grouded theory
  - “Qualitative meta-analysis”
Variables and validity

- **Variables**
  - example with paper review, code inspections, cooker
  - **Independent** (manipulated, input)
  - **Dependent** (output)
  - **Confounding factors** – interference factors bringing threat to internal validity

- **Validity** [7]
  - **Internal** – changes in the dependent variables can be safely attributed to changes in the independent variables
  - **External** – results generalize to setting outside the study
  - **Construct** – variables accurately model the abstract hypothesis
Level (unit) of analysis [1]

- Relates to population and sampling
  - Society
  - Profession
  - External business context
  - Organizational context
  - Project
  - Group/team
  - Individual
  - System
  - Computing element – program, component, algorithm
  - Abstract concept
Data collection

- Definition of **measures** (entity, attribute, unit, counting rules)
- **Iterrater** for subjective measures
- **Quality control** method (completeness and accuracy)
- **Response** rate, representativeness, impact of nonresponse (surveys)
- Data about **drop-outs**, other **performance measures** affected by the treatment (observational)
Data analysis

- Procedures used to control for multiple sampling
- Blind analysis
- Sensitivity analysis
- Data must not violate the assumptions of the tests
- Quality control procedures for results verification
Qualitative vs. Qualitative data/analysis

- Study **can include both**
- Strictly qualitative does not appear in SE

**Quantitative**
- **Numerical** values / attributes
- Numerical expression of **relations or comparison** of alternatives
- Usually examines **causal relations** (answeres why?)

**Qualitative**
- **Non-numerical** data
- Participant observation, document history analysis, interviews
- Related to behavioral, individual and team psychology and sociology
- Usually examines **processes** (answeres how?)
- In the end data translated to numbers (usually) and **combined with quantitative** data and analysis
Results presentation

- Description of statistical procedures
- Statistical package
- Quantitative results – significance levels, magnitude of effects and confidence limits
- Raw data – present/confirm availability
- Descriptive statistics
- Graphical representation
Results interpretation, drawing conclusions

- Statistical **significance** vs. practical **importance**
- Definition of study type (sometimes can not be predicted at the top)
- **Limitations** of the study and the conclusions
  - What do we **assume**? (What we cannot read from the data?) What effect do potential **biases** have? Are the results sufficient to draw these conclusions? Why (not)?
- Given the limits, what does the **data really say**?
  - Are there **ambiguities** in our interpretation? Can the results be **interpreted another way** than we do (need) it? Are our results **believable**?
- **Tie** the results **to the hypothesis**
- How can the result and conclusions be **used by us or others**?
  - Usage in practice
  - Replication
  - Possible generalizations (to what extend/population)
- **What else** can/needs to be done?
Faceted paper classification scheme

[1]

- Topic
- Research approach
- Research method
- Reference field
- Level (unit) of analysis

Useful for generating paper keywords -> better classification and relevance decision on the part of readers

Conclusion of analysis of 369 papers over 5-year period (1995-1999) – SE research is diverse in topic, narrow in research approach and method, inwardly focused from the viewpoint of reference discipline, and technically focused (as opposed to behaviorally focused) in level of analysis
Recommended literature

The Context of the Software Engineering Research

History, goals, top organization, experts and publications, problems and their solutions in SE research
Goals of SE research

- The **main goal** of the research is to raise **effectivity** and **support decision making** in field practices.
- The **outcomes** of the research are **used by**: readers, reviewers, authors, other researchers, meta-analytists, editorial and conference committees, **practitioners**.
- The point of every study is (or should be) the exploration or description of the **relations between** representatives of four “archetype classes” – Actor, Technology, Activity, Software System [10].
History of SE research [1]

- **Software engineering**
  - History is **traceable to** the early **50s**
  - Its first **conferences** were being held in late **60s**
  - Its academic presence did not **begin to separate off** from computer science until early **80s**

- **SE research**
  - **50s** – mainly **ad hoc** research and very **few outlets** to present it, many **findings made** but **not described** in literature for 10 years afterwards
  - The **traceable history** of the field dates to **late 60s**
  - **Qualitative research** had not really appeared until **10-15 years ago**
Top research organizations and lead experts I.

- **University of New South Wales**, Sydney -> National Information Communications Technology Australia (NICTA)
  - The Software Systems Research Group (SSRG)
- **Lund University**, Sweden
  - Software Engineering Research Group (SERG)
  - Claes Wohlin, Martin Höst, Per Runeson
- **Simula Research Laboratory**, Lysaker, Norway
  - Software Engineering
  - Dag I. K. Sjöberg, Tore Dyba, Magne Jörgensen
- **University of Maryland**, USA
  - Victor F. Basili
- **Microsoft**
  - Empirical Software Engineering Group (ESE)
  - Research in Software Engineering (RiSE)
  - Software Analytics
Top research organizations and lead experts II.

- **Keele University**, UK
  - Software engineering Group
  - Barbara A. *Kitchenham*

- **Carnegie Mellon University**, Pittsburgh
  - School of Computer Science (SCS)
  - Computer Science Department (CSD)
  - Institute for Software Research (ISR)
  - Software Engineering Institute (SEI)
  - Mary *Shaw*

- **Chalmers**, University of Ghotenburg
  - Formal Methods in Software Engineering (group)

- Auckland, Wien, Kaiserslautern, Mississippi State University, Delft, Malta, Madrid, USC, Toronto, Alabama, Trondheim (*SINTEF*), Dartmouth College (Shari Lawrence *Pfleeger*), Karlsruhe (Walter F. *Tichy*).
Top journals [1]

- Elsevier’s Information and Software Technology
- Elsevier’s Journal of Systems and Software
- Wiley’s Software Practice and Experience
- IEEE Software
- ACM Transactions on Software Engineering and Methodology
- IEEE Transactions on Software Engineering
Problems in SE research I.

- Absence of unified and universally recognized guidelines, methodologies, taxonomy, terminology
- Little support in universally recognized authority
  - Cochrane Collaboration, Cochrane Database of Systematic Reviews – medicin
  - Human Genome Project – genetics
  - SWEBOK
- Low volume of research
- Gap between practices and academia (research community)
- Purposeless research
- Using students in research experiments and surveys
  - Easy, available, cheap
  - Little or no trust among practitioners
- Lack of (or low quality) results validation and evaluation
- Misused statistical methods
Lack of qualitative research
Incorrect drawing of conclusions, fishing for results, ignoring or throwing away of negative results
Input data (often from industry) not shared along with the paper
General problems with low quality, relevance (significance of results) and usability/impact (research not interesting for industry)
Resources – little financial support from governments, industry and other entities (initiatives, foundations)
Little or none theory building
“Assimilation gap” – 9-15 years between first acquisition of a new technology and its 25% penetration into software development organizations [1]
Suggested solutions to the problems

- Patterning SE after other disciplines (medicine, psychology, sociology, information systems, computer science)
- Including statisticians and practitioners (resources issue)
- Long-term focus on the particular topics, families of studies (replication, similar studies, generalization of results)
- Community-wide cooperation towards stable and recognized terminology, taxonomy, guideline creation, theory building and authority foundation to ensure higher research quality and effectivity
- Initiatives and organization bringing resources
- Insightful questions
- Involving others (building relationships)
- Education (SE research courses)
Conclusions

- SE research as a field has many different problems
- Particular studies also struggle with insufficient quality, relevance and impact
- Suggestions exist and steps are being taken towards minimalization of the issues impact but its a long-term and slow process
- General advice includes
  - Rigorous study design and description
  - Including statisticians and as much of practitioners (or at least industry data) participation as possible
  - Education of researchers
  - Establishing authorities and financial support system
  - Using othes sciences as patterns
Thank You

Ask me whatever you want, I'll answer whatever I can
References I.

References II.

Research in Software Engineering: An Analysis of the Literature

By
## Findings for computing topic

<table>
<thead>
<tr>
<th>1.0</th>
<th>Problem-solving concepts</th>
<th>5.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Algorithms</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>1.2</td>
<td>Mathematics/computational science</td>
<td>0%</td>
</tr>
<tr>
<td>1.3</td>
<td>Methodologies (object, function/process, information/data, event, business rules, …)</td>
<td>4.9%</td>
</tr>
<tr>
<td>1.4</td>
<td>Artificial intelligence</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>2.0</td>
<td>Computer concepts</td>
<td>10.9%</td>
</tr>
<tr>
<td>2.1</td>
<td>Computer/hardware principles/architecture</td>
<td>0%</td>
</tr>
<tr>
<td>2.2</td>
<td>Inter-computer communication (networks, distributed systems)</td>
<td>9.5%</td>
</tr>
<tr>
<td>2.3</td>
<td>Operating systems (as an augmentation of hardware)</td>
<td>1.4%</td>
</tr>
<tr>
<td>2.4</td>
<td>Machine/assembler-level data/instructions</td>
<td>0%</td>
</tr>
<tr>
<td>3.0</td>
<td>Systems/software concepts</td>
<td>54.8%</td>
</tr>
<tr>
<td>3.1</td>
<td>System architecture/engineering</td>
<td>1.9%</td>
</tr>
<tr>
<td>3.2</td>
<td>Software life-cycle/engineering (incl. requirements, design, coding, testing, maintenance)</td>
<td>8.7%</td>
</tr>
<tr>
<td>3.3</td>
<td>Programming languages</td>
<td>3.8%</td>
</tr>
<tr>
<td>3.4</td>
<td>Methods/techniques (incl. reuse, patterns, parallel processing, process models, data models, etc)</td>
<td>18.2%</td>
</tr>
<tr>
<td>3.5</td>
<td>Tools (incl. compilers, debuggers)</td>
<td>12.2%</td>
</tr>
<tr>
<td>3.6</td>
<td>Product quality (incl. performance, fault tolerance)</td>
<td>8.4%</td>
</tr>
<tr>
<td>3.7</td>
<td>Human–computer interaction</td>
<td>1.1%</td>
</tr>
<tr>
<td>3.8</td>
<td>System security</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>4.0</td>
<td>Data/information concepts</td>
<td>7.6%</td>
</tr>
<tr>
<td>4.1</td>
<td>Data/file structures</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>4.2</td>
<td>Data base/warehouse/mart organization</td>
<td>4.6%</td>
</tr>
<tr>
<td>4.3</td>
<td>Information retrieval</td>
<td>1.4%</td>
</tr>
<tr>
<td>4.4</td>
<td>Data analysis</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>4.5</td>
<td>Data security</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>5.0</td>
<td>Problem-domain-specific concepts (use as a secondary subject, if applicable, or as a primary subject if there is no other choice)</td>
<td>2.7%</td>
</tr>
<tr>
<td>5.1</td>
<td>Scientific/engineering (incl. bio-informatics)</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>5.2</td>
<td>Information systems (incl. decision support, group support systems, expert systems)</td>
<td>1.6%</td>
</tr>
<tr>
<td>5.3</td>
<td>Systems programming</td>
<td>0%</td>
</tr>
<tr>
<td>5.4</td>
<td>Real-time (incl. robotics)</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>5.5</td>
<td>Edutainment (incl. graphics)</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>6.0</td>
<td>Systems/software management concepts</td>
<td>11.6%</td>
</tr>
<tr>
<td>6.1</td>
<td>Project/product management (incl. risk management)</td>
<td>3.3%</td>
</tr>
<tr>
<td>6.2</td>
<td>Process management</td>
<td>2.2%</td>
</tr>
<tr>
<td>6.3</td>
<td>Measurement/metrics (development and use)</td>
<td>6.2%</td>
</tr>
<tr>
<td>6.4</td>
<td>Personnel issues</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>7.0</td>
<td>Organizational concepts</td>
<td>1.9%</td>
</tr>
<tr>
<td>7.1</td>
<td>Organizational structure</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>7.2</td>
<td>Strategy</td>
<td>0%</td>
</tr>
<tr>
<td>7.3</td>
<td>Alignment (incl. business process reengineering)</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>7.4</td>
<td>Organizational learning/knowledge management</td>
<td>0%</td>
</tr>
<tr>
<td>7.5</td>
<td>Technology transfer (incl. innovation, acceptance, adoption, diffusion)</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>7.6</td>
<td>Change management</td>
<td>0%</td>
</tr>
<tr>
<td>7.7</td>
<td>Information technology implementation</td>
<td>0%</td>
</tr>
<tr>
<td>7.8</td>
<td>Information technology usage/operation</td>
<td>0%</td>
</tr>
<tr>
<td>7.9</td>
<td>Management of ‘computing’ function</td>
<td>0%</td>
</tr>
<tr>
<td>7.11</td>
<td>IT Impact</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>7.11</td>
<td>Computing/information as a business</td>
<td>0%</td>
</tr>
<tr>
<td>7.12</td>
<td>Legal/ethical/cultural/political (organizational) implications</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>8.0</td>
<td>Societal concepts</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>8.1</td>
<td>Cultural implications</td>
<td>0%</td>
</tr>
<tr>
<td>8.2</td>
<td>Legal implications</td>
<td>0%</td>
</tr>
<tr>
<td>8.3</td>
<td>Ethical implications</td>
<td>0%</td>
</tr>
<tr>
<td>8.4</td>
<td>Political implications</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>9.0</td>
<td>Disciplinary issues</td>
<td>3.5%</td>
</tr>
<tr>
<td>9.1</td>
<td>‘Computing’ research</td>
<td>1.1%</td>
</tr>
<tr>
<td>9.2</td>
<td>‘Computing’ curriculum/teaching</td>
<td>2.4%</td>
</tr>
</tbody>
</table>
Findings for research approach

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive:</td>
<td></td>
<td>27.9%</td>
</tr>
<tr>
<td>DS</td>
<td>Descriptive system</td>
<td>8.1%</td>
</tr>
<tr>
<td>DO</td>
<td>Descriptive other</td>
<td>18.2%</td>
</tr>
<tr>
<td>DR</td>
<td>Review of literature</td>
<td>1.6%</td>
</tr>
<tr>
<td>Evaluative:</td>
<td></td>
<td>13.8%</td>
</tr>
<tr>
<td>ED</td>
<td>Evaluative—deductive</td>
<td>4.3%</td>
</tr>
<tr>
<td>EI</td>
<td>Evaluative—interpretive</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>EC</td>
<td>Evaluative—critical</td>
<td>1.4%</td>
</tr>
<tr>
<td>EO</td>
<td>Evaluative—other</td>
<td>7.3%</td>
</tr>
<tr>
<td>Formulative:</td>
<td></td>
<td>55.3%</td>
</tr>
<tr>
<td>FF</td>
<td>Formulative-framework</td>
<td>4.1%</td>
</tr>
<tr>
<td>FG</td>
<td>Formulative-guidelines/standards</td>
<td>4.3%</td>
</tr>
<tr>
<td>FM</td>
<td>Formulative-model</td>
<td>9.8%</td>
</tr>
<tr>
<td>FP</td>
<td>Formulative—process, method, algorithm</td>
<td>36.0%</td>
</tr>
<tr>
<td>FT</td>
<td>Formulative-classification/taxonomy</td>
<td>1.1%</td>
</tr>
<tr>
<td>FC</td>
<td>Formulative-concept</td>
<td>3.0%</td>
</tr>
</tbody>
</table>
## Findings for research method

<table>
<thead>
<tr>
<th>Code</th>
<th>Research Method</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>Action research</td>
<td>0%</td>
</tr>
<tr>
<td>CA</td>
<td>Conceptual analysis</td>
<td>43.5%</td>
</tr>
<tr>
<td>CAM</td>
<td>Conceptual analysis/mathematical</td>
<td>10.6%</td>
</tr>
<tr>
<td>CI</td>
<td>Concept implementation (proof of concept)</td>
<td>17.1%</td>
</tr>
<tr>
<td>CS</td>
<td>Case study</td>
<td>2.2%</td>
</tr>
<tr>
<td>DA</td>
<td>Data analysis</td>
<td>2.2%</td>
</tr>
<tr>
<td>DI</td>
<td>Discourse analysis</td>
<td>0%</td>
</tr>
<tr>
<td>ET</td>
<td>Ethnography</td>
<td>0%</td>
</tr>
<tr>
<td>FE</td>
<td>Field experiment</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>FS</td>
<td>Field study</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>GT</td>
<td>Grounded theory</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>HE</td>
<td>Hermeneutics</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>ID</td>
<td>Instrument development</td>
<td>0%</td>
</tr>
<tr>
<td>LH</td>
<td>Laboratory experiment (human subjects)</td>
<td>3.0%</td>
</tr>
<tr>
<td>LR</td>
<td>Literature review/analysis</td>
<td>1.1%</td>
</tr>
<tr>
<td>LS</td>
<td>Laboratory experiment (software)</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>MA</td>
<td>Meta-analysis</td>
<td>0%</td>
</tr>
<tr>
<td>MP</td>
<td>Mathematical proof</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>PA</td>
<td>Protocol analysis</td>
<td>0%</td>
</tr>
<tr>
<td>PH</td>
<td>Phenomenology</td>
<td>0%</td>
</tr>
<tr>
<td>SI</td>
<td>Simulation</td>
<td>1.1%</td>
</tr>
<tr>
<td>SU</td>
<td>Descriptive/exploratory survey</td>
<td>1.6%</td>
</tr>
<tr>
<td>Code</td>
<td>Field</td>
<td>Percentage</td>
</tr>
<tr>
<td>------</td>
<td>--------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>CP</td>
<td>Cognitive psychology</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>SB</td>
<td>Social and behavioral science</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>CS</td>
<td>Computer science</td>
<td>0%</td>
</tr>
<tr>
<td>SC</td>
<td>Science</td>
<td>0%</td>
</tr>
<tr>
<td>EN</td>
<td>Engineering</td>
<td>0%</td>
</tr>
<tr>
<td>EC</td>
<td>Economics</td>
<td>0%</td>
</tr>
<tr>
<td>LS</td>
<td>Library science</td>
<td>0%</td>
</tr>
<tr>
<td>MG</td>
<td>Management</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>MS</td>
<td>Management science</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>PA</td>
<td>Public administration</td>
<td>0%</td>
</tr>
<tr>
<td>PS</td>
<td>Political science</td>
<td>0%</td>
</tr>
<tr>
<td>OT</td>
<td>Other</td>
<td>98.1%</td>
</tr>
<tr>
<td>Code</td>
<td>Description</td>
<td>Percentage</td>
</tr>
<tr>
<td>------</td>
<td>-------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>SOC</td>
<td>Society</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>PRO</td>
<td>Profession</td>
<td>2.4%</td>
</tr>
<tr>
<td>EXT</td>
<td>External business context</td>
<td>0%</td>
</tr>
<tr>
<td>OC</td>
<td>Organizational context</td>
<td>2.2%</td>
</tr>
<tr>
<td>PR</td>
<td>Project</td>
<td>4.1%</td>
</tr>
<tr>
<td>GP</td>
<td>Group/team</td>
<td>1.4%</td>
</tr>
<tr>
<td>IN</td>
<td>Individual</td>
<td>1.4%</td>
</tr>
<tr>
<td>CS</td>
<td>System</td>
<td>10.6%</td>
</tr>
<tr>
<td>CE</td>
<td>Computing element—program, component, algorithm</td>
<td>27.9%</td>
</tr>
<tr>
<td>AC</td>
<td>Abstract concept</td>
<td>49.9%</td>
</tr>
</tbody>
</table>
Preliminary Guidelines for Empirical Research in Software Engineering

By
Context Guidelines

C1) Be sure to **specify** as much of the **industrial context** as possible. In particular, clearly define the entities, attributes, and measures that are capturing the contextual information.

C2) **If a specific hypothesis is being tested, state it** clearly prior to performing the study and **discuss the theory** from which it is derived, so that its implications are apparent.

C3) **If the research is exploratory, state** clearly and, prior to data analysis, what **questions** the investigation is intended to **address and how** it will address them.

C4) **Describe research** that is **similar to**, or has a **bearing on**, the current research and **how current work relates to it**.
D1) Identify the population from which the subjects and the objects are drawn.

D2) Define the process by which the subjects and objects were selected.

D3) Define the process by which subjects and objects are assigned to treatment.

D4) Restrict yourself to simple study designs or, at least, to designs that are fully analyzed in the statistical literature. If you are not using a well-documented design and analysis method, you should consult a statistician to see whether yours is the most effective design for what you want to accomplish.

D5) Define the experimental unit.
D6) **For formal experiments**, perform a *pre-experiment or precalculation* to identify or *estimate* the minimum required *sample size*.

D7) **Use** appropriate levels of *blinding*.

D8) If you cannot avoid evaluating your own work, then **make explicit** any **vested interest** (including your sources of support) and report *what* you have done to minimize bias.

D9) **Avoid** the use of *controls unless* you are sure the control situation *can be unambiguously defined*.

D10) **Fully define** all *treatments* (interventions).

D11) **Justify** the choice of *outcome measures* in terms of their relevance to the objectives of the empirical study.
Data Collection Guidelines

DC1) Define all software measures fully, including the entity, attribute, unit and counting rules.

DC2) For subjective measures, present a measure of interrater agreement, such as the kappa statistic or the intraclass correlation coefficient for continuous measures.

DC3) Describe any quality control method used to ensure completeness and accuracy of data collection.

DC4) For surveys, monitor and report the response rate and discuss the representativeness of the response and the impact of nonresponse.

DC5) For observational studies and experiments, record data about subjects who drop out form the studies.

DC6) For observational studies and experiments, record data about other performance measures that may be affected by the treatment, even if they are not the main focus of the study.
Analysis Guidelines

A1) Specify any procedures used to control for multiple testing.
A2) Consider using blind analysis.
A3) Perform sensitivity analyses.
A4) Ensure that the data do not violate the assumptions of the tests used on them.
A5) Apply appropriate quality control procedures to verify your results.
**Presentation Guidelines**

P1) Describe or cite a reference for all statistical procedures used.

P2) Report the statistical package used.

P3) Present quantitative results as well as significance levels. Quantitative results should show the magnitude of effects and the confidence limits.

P4) Present the raw data whenever possible. Otherwise, confirm that they are available for confidential review by the reviewers and independent auditors.

P5) Provide appropriate descriptive statistics.

P6) Make appropriate use of graphics.
Interpretation guidelines

1) Define the population to which inferential statistics and predictive models apply.
2) Differentiate between statistical significance and practical importance.
3) Define type of study.
4) Specify any limitations of the study.
What Makes Good Research in Software Engineering?

By
Another classification scheme of research papers

- **Type of question**
  - Method or means of development
  - Method for analysis
  - Design, evaluation or analysis of a particular instance
  - Generalization or characterization
  - Feasibility

- **Type of result**
  - Procedure or technique
  - Qualitative or descriptive model

- **Empirical model**
- **Analytic model**
- **Notation or tool**
- **Specific solution**
- **Answer or judgment**
- **Report**

- **Type of validation**
  - Analysis
  - Experience
  - Example
  - Evaluation
  - Persuasion
  - Blatan Assertion
Experimental Validation in Software Engineering

By
Another research paper classification

- Replicated
- Synthetic
- Dynamic analysis
- Simulation
- Project monitoring
- Case study
- Assertion
- Field study
- Literature search
- Legacy data
- Lessons learned
- Static analysis
Misleading Metrics and Unsound Analyses

By
Misuse of metrics in IBM Australia I.

- International standart ISO/IEC 15393 gives inapproriate advice for measuring SE processes
- When combined with the CMM/CMMI level 4 requirement for statistical process control can encourage the use of misleading metrics and the adoption of inappropriate data aggregation and analysis technique
- Applying normal distribution methods to non-normal data
- **Instability** of the productivity metrics
- Distribution of two unrelated variables can be very unstable if the values are small
- 13 – high productivity in both
- 2 – high productivity in 1st, but very small (negligible importance)
- 1 and 4 – relatively productive but relatively small
- 3 – extremely low productivity in 2nd, but in 1st doesn’t stand out (simple productivity plot compresses low productivity values at the bottom)