[simula . research laboratory]

Collection and evaluation of results from research studies and practice-based experience

Learning goals: Improved ability to collect and evaluate research studies and experience.

Supporting texts:

- Guidelines for performing Systematic Literature Reviews in Software Engineering, Kitchenham et al, 2007
- Example: Forecasting of Software Development Work Effort: Evidence on Expert Judgment and Formal Models, Jørgensen, 2007 (International Journal of Forecasting)

Main Steps

- Specifying the research question(s)
- Developing a review protocol
- Identification of relevant studies and experience
- Evaluation of studies and experience
- Synthesis of studies and experience

[simula . research laboratory]

Example: Research Questions

- Should we expect more accurate effort estimates when applying expert judgment or models?
- When should software development effort estimates be based on expert judgment, on models, or on a combination of expert judgment and models?

Example: Review Protocol

- The identification of relevant studies is based on an manual examination of papers in potentially relevant journals and a search in the library database Inspec for papers including the terms (‘*effort estimation*’ OR ‘*cost estimation*’) AND ‘*software development*’
- Last search conducted February 2006

Example: Identification

- In total, seventeen relevant papers were identified. One of the papers was excluded due to incomplete information about how the estimates were derived, which left sixteen papers for review.
- I did not search for practice-based experience, since I had identified a sufficiently high number of relevant published studies. If I had to collect practice-based experience, this is how I probably would have done that:
  - Find organizations that use both models and expert estimation
  - Ask them about their experience with both, with the control questions:
    - Measurement of performance, or opinion-based experience?
    - Used on same type of projects?
    - Used at the same stage of the estimation process?
    - Proper use of models?
    - Proper use of methods?
  - (My experience is that the same review protocols and checklists are applicable for evaluation of published studies and experience, see next slide.)

Example: Evaluation

- Sixteen studies are reviewed with respect to important contextual factors, i.e., the factors identified in the discussion in Sections 2 and 3. The main design factors and results reviewed for each study are as follows:
- DESIGN FACTORS
  - Study design
  - Estimation method selection process
  - Estimation models
  - Calibration level
  - Model use expertise and degree of mechanical use of model
  - Expert judgment process
  - Expert judgment estimation expertise
  - Possible motivational biases in estimation situation
  - Fairness limitations
- RESULTS:
  - Accuracy

Example: Synthesis

- A comparison of the average accuracy of the models with the average accuracy of the experts shows that ten studies found increased accuracy with the use of expert judgment and six with the use of estimation models.
- Based on the modest evidence to date, two conditions for producing more accurate expert judgment-based effort estimates seem to be that the models are not calibrated to the organization using them, and that the experts possess important contextual information not included in the formal models and apply it efficiently.
- The use of models, either alone or in combination with expert judgment, may be particularly useful when i) there are situational biases that are believed to lead to a strong bias towards overoptimism; ii) the amount of contextual information possessed by the experts is low; and iii) the models are calibrated to the organization using them.