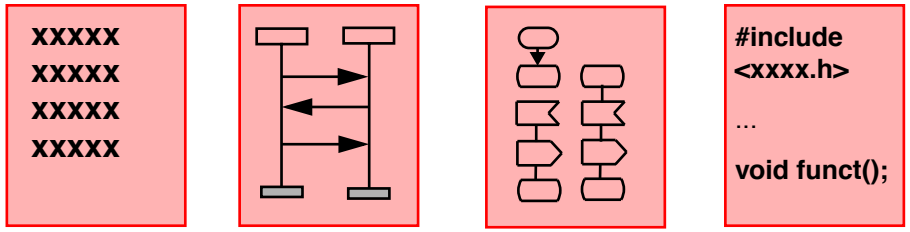
Measuring the Effect of Formalization

Ketil Stølen, Peter Mohn*

OECD Halden Reactor Project
Institute for Energy Technology
Halden, Norway



* On leave from the Swiss Federal Nuclear Safety Inspectorate

Lack of Experimentation

“ A significant segment of the software industry converted from C to C++ at a substantial cost in retaining. ... I’m not aware of any solid evidence showing that C++ is superior to C with respect to programmer productivity or software quality. ”

Tichy, IEEE Computer
May, 1998

Lack of Experimentation

"... , functional programming, object-oriented programming, and formal methods are all thought to improve programmer productivity, program quality, or both. It is surprising that none of these obviously important claims have ever been tested systematically, ..."

Tichy, IEEE Computer
May, 1998

Arguments against Experimentation

- Traditional scientific method does not apply
- Demonstrations will suffice
- Technology changes too fast
- Too many variables
- Too costly
- Progress will slow
- You will never get it published
- Current practise good enough

Study by Tichy et al

A study based on a random sample of the papers ACM published in 1993 found that

- 40 percent of the papers with claims that needed empirical support had none at all
- In software related journals this fraction was 50 percent
- In the non-computer-science journal, Optical Engineering, this fraction was 15 percent

Study by Zelkowitz/Wallace

A study based on 612 papers seems to confirm the results of Tichy et al

- **About 40 percent of the papers with claims that needed empirical support had none at all**
- **Similar studies in physics, psychology, and anthropology carried out by Zelkowitz/Wallace found much smaller percentages of unvalidated papers**

There are Problems, but they can be Overcome

- Experimentation is difficult
  - many potential flaws
  - unrealistic assumptions
  - data manipulation
  - fraud
- But the situation is not worse than in several other fields
  - testing of pharmaceuticals
  - social sciences

OECD Halden Reactor Project (HRP)

- International cooperative effort
- 40 years anniversary
- Run by the OECD Nuclear Energy Agency
- Hosted by the Institute for Energy Technology, Halden
- 20 member countries
- 100 nuclear organizations
- Safety in the design and operation of installations
- Two main parts: fuel and control rooms

Formal Methods at the HRP

- Research on formal methods has some tradition
  - previously specialized towards safety critical systems and formal development
  - recently the scope has been widened
- HRP organizations suggested several directions
  - integration in real system developments
  - combining formal and informal techniques
  - training of personnel
  - measuring the effect of formalization

Formal Methods at the HRP

- Area of interest
  - interaction, communication, distribution
  - real-time, object-orientation
  - Statecharts, SDL, MSC, UML
  - state-of-the-art CASE-tools
- Preliminary results
  - HWR-522 summarizes experiences
  - HWR-523 evaluates 11 specification languages
  - HWR-526 surveys model checking and theorem proving

Starting Point

- Two papers directed towards VDM/B
  - J. Bicarregui, J. Dick, E. Woods

    Quantitative analysis of an application of formal methods

    Formal Methods Europe (FME) 1996
  - J. Draper, H. Treharne, T. Boyce, B. Ormsby

    Evaluating the B-method on an avionics example

    Data Systems in Aerospace (DAISA) 1996

Set Up of Experiment

- Metrics to compare with conventional techniques
  - faults per 1000 lines of code found by unit and integration tests
  - faults per 1000 lines of code found by validation test
  - faults per 1000 lines of code found by customers
  - person months of effort per 1000 lines of code
- To compare the relative effectiveness of various stages
  - faults are registered throughout the development
  - at which stage/activity the fault was introduced
  - at which stage/activity the fault was discovered

Set Up of Experiment

- The notion of fault is defined as follows:
  - a fault is found when a change is required to a design decision made at an earlier development stage
  - a design made and corrected within the same stage is not considered as a fault
- The developments were based on a waterfall process

Experiences

- Systems engineers
  - motivation
  - executive support
- Lines of code
  - code-generation
  - estimates
- Faults
  - room for interpretation
  - satisfaction

Experiences

- Requirements
  - glass-box
  - black-box
- Software process
  - overlapping stages
  - iterative process
  - component based
- Metrics

Satisfaction

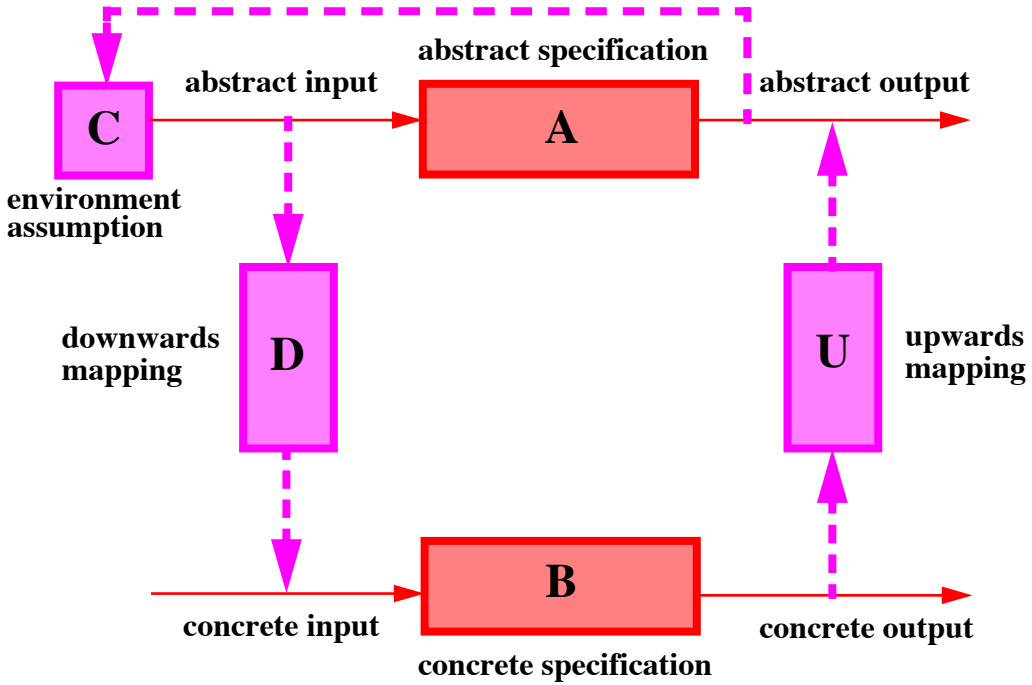
- Has been studied in the context of formal development methods --- often referred to as refinement
- Three notions of refinement
  - property refinement

    supports requirements capture and step-wise development
  - interface refinement

    supports different levels of abstraction, reuse and adaptation
  - conditional refinement

    supports the introduction of boundedness constraints and synchronization

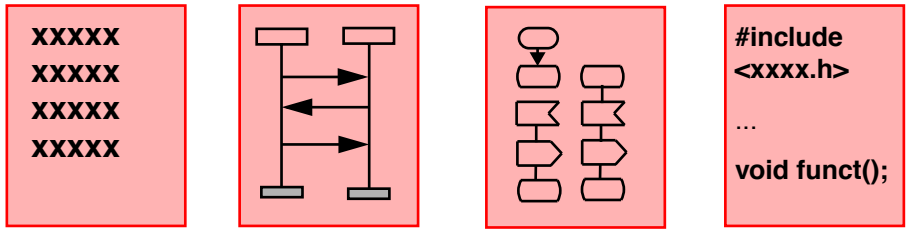
Refinement in FOCUS



MSC in the Context of FOCUS

- In the MSC literature we find statements like
  - a MSC document is never absolutely finished
  - a MSC describes a set of example runs
  - a MSC document does not define the complete set of runs
- Two interpretations
  - a MSC document describes runs that not necessarily are potential runs of the final implementation
  - any run allowed by a MSC document must be a potential run of the final implementation
- Neither fits with FOCUS

Summary



- Difficult, but important
- We may learn from other fields
- Preliminary experiment identified problems
- We look forward to feedback on how they should be tackled

Future Plans

- Recommendations for how to measure the effect of formalization will be worked out
  - tested in system developments at the HRP
  - data from these experiments will be analysed
- We will try to create cooperation on this issue
  - realistic experiments
  - workshop
  - projects

Email:Ketil.Stoelen@hrp.no