Empirical validation research methods

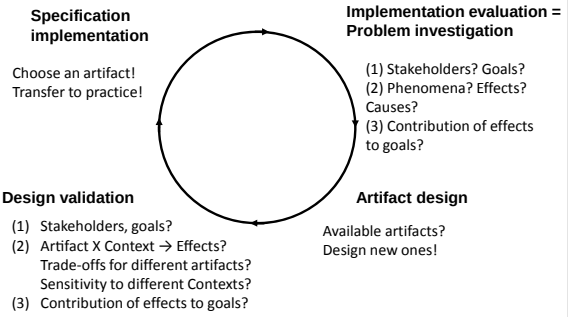
Roel Wieringa
University of Twente
The Netherlands

Design science

- Design and investigation of artifacts
- Questions:
  - What artefact(s) are you developing?
  - How will you investigate its properties?

1. The engineering cycle: developing artifacts
2. The technical validation problem
3. Technical validation research questions
4. Validation research methods
5. Examples

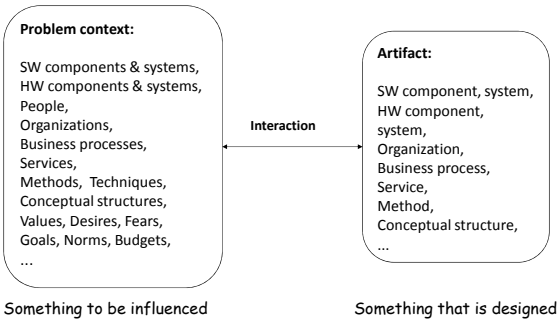
The engineering cycle



Examples of problems (research goals)

- Developing an extension of i*
- Developing a genome ontology
- Developing an argumentation-based risk assessment technique
- ….
- Questions:
  - What problem are you solving?
  - For whom? Stakeholders.
  - What are current their experiences? Phenomena.
  - Goals? Contribution?

Problems and artifacts



Examples of treatments

- Developing a new technique to link business goals to enterprise architecture
- Developing new techniques for agile requiremens prioritization
- Developing an algorithm for P2P document lookup with a distributed hash table
- Developing a directional antenna for TV reception in cars
- Developing techniques to deal with interference in wireless city networks
- *What is the problem context in these cases?*

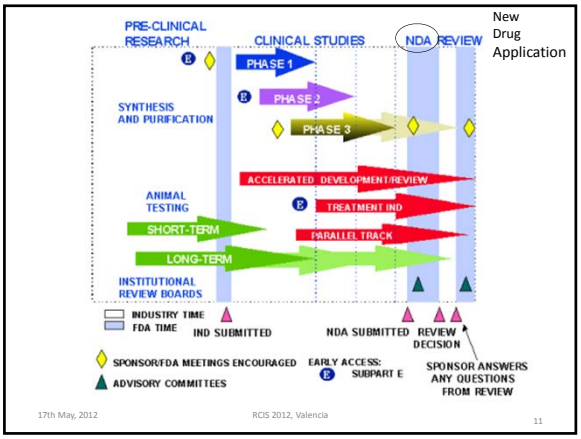
1. The engineering cycle
2. The technical validation problem
3. Technical validation research questions
4. Validation research methods
5. Examples

The fundamental problem of validation

- The artifact has not been implemented (transferred to practical problem contexts) yet.
  - It is not interacting with its intended problem context yet
  - You cannot have it interact with the intended problem context
- So how to validate it?

Example: drug validation research

- Drug = Medicine = Artifact
- Patient's body = Problem context
- Guidelines for validating medicine have been published by the U.S. Food and Drug Administration
- http://www.fda.gov/cder/handbook/develop.htm
- New drug development process: http://www.fda.gov/Drugs/DevelopmentApprovalProcess/SmallBusinessAssistance/ucm053131.htm
- See also http://en.wikipedia.org/wiki/Clinical_research



- Pre-clinical research
  - Synthesis and purification.
    - (1) Investigate human body,
    - (2) Design a new drug and
    - (3) Investigate its properties in the laboratory
  - Treat animals with it in short-term experiments (few weeks)
  - Treat animals with it in long-term experiments (few weeks to several years)

1. Problem
2. Design
3. Validation
4. Validation
5. Validation

Animals are models of patients

- Pre-clinical research
  - Synthesis and purification. (1) Investigate human body, (2) Design a new drug and (3) Investigate its properties in the laboratory
  - Treat animals with it in short-term experiments (few weeks)
  - Treat animals with it in long-term experiments (few weeks to several years)
- Clinical research
  - Phase 1: Treat **healthy people** with it to understand the mechanisms and possible side effects of the drug (20 to 80 volunteers)
  - Phase 2: Treat **ill people** with it to understand mechanisms, side effects and contra-indications (few hundred volunteers)
  - Phase 3: Treat **ill people** with it to investigate effectiveness, side effects and contra-indications on a sufficiently large sample to generalize (few 100 to few 1000 volunteers)

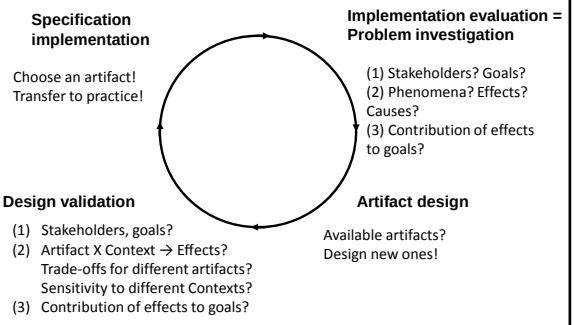
5. Validation
6. Validation
7. Validation

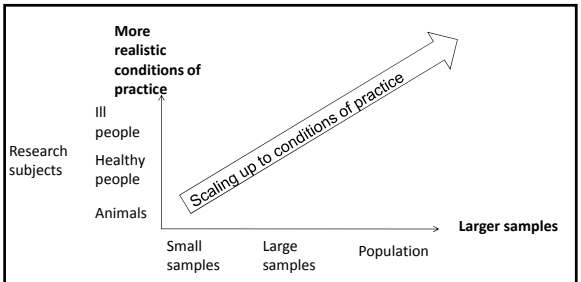
- Side effects,
- Mechanisms,
- Contra-indications

- Pre-clinical research
  - Synthesis and purification. (1) Investigate human body, (2) Design a new drug and (3) Investigate its properties in the laboratory
  - Treat animals with it in short-term experiments (few weeks)
  - Treat animals with it in long-term experiments (few weeks to several years)
- Clinical research
  - Phase 1: Treat healthy people with it to understand the mechanisms and possible side effects of the drug (20 to 80 volunteers)
  - Phase 2: Treat ill people with it to understand mechanisms, side effects and contra-indications (few hundred volunteers)
  - Phase 3: Treat ill people with it to investigate effectiveness, side effects and contra-indications on a sufficiently large sample to generalize (few 100 to few 1000 volunteers)
- Post availability
  - Phase 4: continue clinical studies to better understand the drug
  - Surveillance studies in the field

Evaluation

The engineering cycle





- Once the drug can be produced ("purified") in the lab, its use is scaled up to conditions of practice

Validation is modelling

- Animals, healthy volunteers, and ill volunteers are used as **models** of arbitrary patients
- Conclusions about the models are transferred to arbitrary patients
- The model is a composite system (Problem X Artifact)
- The "X" is the treatment

Validation is theory building

- Development starts with an initial treatment theory (hunch, hope, hypothesis):
  - "Artifact treats problem successfully".
- This theory is developed together with the artifact
  - It makes a cause effect statement (Artifact X Context) → Effects
  - Side effects as well as desirable effects
  - It tries to explain this in terms of mechanisms
  - It claims this to be true of a certain class of contexts

Validation is risk assessment

- Validation is asking what the risk is of designing the wrong artifact
  - Benefit not high enough
  - Unexpected harm
- Generalization of a design theory to the entire class of problems (population) entails a large risk to stakeholders
  - They must commit resources if they will use your artifact
  - *They* have to live with the results

1. The engineering cycle
2. The technical validation problem
3. Technical validation research questions
4. Validation research methods
5. Examples

Justification

- You do not design artifacts but justifications
  - No one is interested in an artifact without justification
  - Your artifact will get lost anyway
  - Your justification is also useful without concrete artifact
- Justifications are needed to justify investing in your artifact
  - What are the expected **benefits** to stakeholders?
  - What are the **costs** for stakeholders?
  - What is the **risk** for stakeholders that your artifact will not mitigate the problem or will make matters worse?

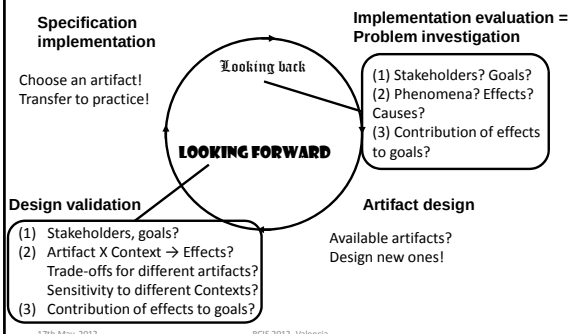
Validation research

- Validation provides the information to justify your design
- There are four standard validation research questions

1. Treatment of Context by Artifact produces Effects?
2. Trade-offs for different artifacts?
3. Sensitivity for different Contexts?
4. Effects contribute to stakeholder goals?

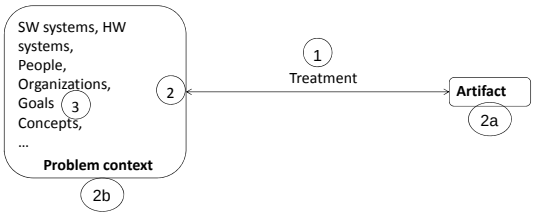
- The answers to these questions give information about **cost** and **benefit** of this artifact compared to others
- Uncertainty about these answers is a **risk**; hence you should make clear how (un)certain your are about them

The engineering cycle



- **Trade-off analysis** is needed to compare your treatment with other treatments
  - No treatment satisfies all design criteria equally well
  - Often only part of an artifact is applied
- **Sensitivity analysis** is needed to generalize to similar problem domains
- They are both forms of generalization

Validation in context



1. Who are the stakeholders and what are their goals?
2. What effects do the treatment have (on context)?
   a) What happens if we change the artifact design?
   b) What happens if we change our assumptions about the context?
3. How does this contribute to stakeholder goals?

Example research reports

- *DOA estimation for satellite TV reception in cars*
  - ***Artifact:*** *DOA estimation algorithm.*
  - ***Context:*** *arriving satellite TV signals, receiver hardware, beamformer and beamsteering systems, car, driver, passengers.*
  - ***Treatment:*** *Reception of signals, determination of direction while car is moving,* ***in order that*** *passengers can watch satellite TV.*
  - ***Validation questions:***
    - *Effect: Can we recognize DOA?*
    - *Trade-off: Comparison of algorithms?*
    - *Sensitivity to car speed, processor properties?*
    - *Contribution: How accurate; space efficient, time efficient?*

- *Information exchange architecture for hospitals and insurance companies*
  - ***Artifact:*** *IT architecture*
  - ***Context:*** *IT of hospitals and insurance companies, business processes, laws and regulations*
  - ***Treatment:*** *Accepting and delivering data both ways,* ***in order that*** *hospitals can get medical activities funded and insurance companies can pay out insurance claims*
  - ***Validation questions:***
    - *Effect: can the required data be exchanged?*
    - *Trade-offs: No comparison done*
    - *Sensitivity analysis: not done*
    - *Contribution: Does it match currently existing IT architecture in hospitals? Does it comply to privacy laws?*

Remark:
•Operationalization of "match" is "adaptation at acceptable cost".

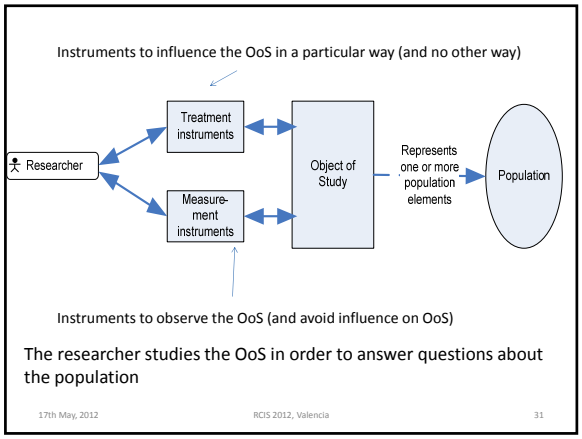
- *Claims-based workflow management*
  - ***Artifact:*** *claims-based WFM system*
  - ***Context:*** *Employees, business processes, access permissions, business rules*
  - ***Treatment:*** *Process requests to perform activities, respecting access permissions and business rules,* ***in order that*** *employee can perform activities in a flexible manner*
  - ***Validation questions*** *(not asked in the thesis):*
    - *Can claim-based WF handle business processes?*
    - *Trade-off: (Not asked) What overhead is caused by flexibility/*
    - *Sensitivity: (Not asked) Does claim-based WF M scale up to realistic numbers of processes?*
    - *Contribution: can CBWF handle the processes flexibly?*

Remarks:
•This is an explorative "can we do it" research
•Flexibility must be operationalized

Discussion

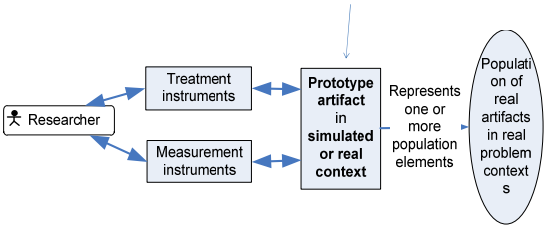
- What are the validation questions in your (planned) validation research?

1. The engineering cycle
2. The technical validation problem
3. Technical validation research questions
4. Validation research methods
5. Examples



The researcher studies the OoS in order to answer questions about the population

In empirical validation research,
- The artifact must be prototyped
- The context must be simulated



Empirical validation research methods

- Illustration using a small example (not a research method)
- Expert opinion (**opinion survey**)
- Demonstration using a realistic example (**lab demo**)
- Validation using a prototype and simulated context
  - **Parallel group lab experiment**
  - **Lab test**
- Validation using a prototype in a real problem
  - **Parallel group field experiment**
  - **Field test** (technical action research)
- No observational case study: No real-world cases to observe yet

Illustration

- **Illustration using a small example (not a research method)**
  - **Researcher applies uses artifact to solve a toy problem**
  - **Purpose is to explain the artifact and its intended treatment**
  - **Example is easy to understand**
  - **Examples fit in a research paper that explains how the artifact is designed**
- Expert opinion (**opinion survey**)
- Demonstration using a realistic example (**lab demo**)
- Validation using a prototype and simulated context
  - **Parallel group lab experiment**
  - **Lab test**
- Validation using a prototype in a real problem
  - **Parallel group field experiment**
  - **Field test** (technical action research)

(Non-empirical methods)

- Artefact is formally specified (e.g. an algorithm)
- Formal model of context
- Formal / mathematical proof of Artifact & Context → Effects
- If artefact is a method, the mathematics is usually trivial
  - Method: *Pre-wash; main wash; rinse, spin-dry; hang to dry; iron.*
  - Formal verification: *With suitable formalization you can prove that when the input is clothes, the output is clean and ironed clothes.*
- In any case, formal verification is never enough
  - It assumes a model of the context
  - What will happen in a real context?
  - Conditions of practice: *blurring colors, wrong labels, ….*

Opinion survey

- Illustration using a small example (not a research method)
- Expert opinion (**opinion survey**)
  - **Researcher asks practitioners about perceived usability and utility of new artifact in the contexts that they know first-hand**
  - **Interview and/or**
  - **Questionnaire and/or**
  - **Focus group**
- Demonstration using a realistic example (**lab demo**)
- Validation using a prototype and simulated context
  - **Parallel group lab experiment**
  - **Lab test**
- Validation using a prototype in a real problem
  - **Parallel group field experiment**
  - **Field test** (technical action research)

Lab demo

- Illustration using a small example (not a research method)
- Expert opinion (**opinion survey**)
- **Demonstration using a realistic example (lab demo)**
  - **Researcher applies the artifact in the lab to a realistic problem: past real-world project, or a research benchmark**
  - **Purpose is to justify that the artifact *could* be used in practice and/or to compare its performance on this benchmark with that of others**
  - **Explanation in terms of mechanisms**
  - **No realistic environment**
- Validation using a prototype and simulated context
  - **Parallel group lab experiment**
  - **Lab test**
- Validation using a prototype in a real problem
  - **Parallel group field experiment**
  - **Field test** (technical action research)

Parallel groups lab experiment

- Illustration using a small example (not a research method)
- Expert opinion (**opinion survey**)
- Demonstration using a realistic example (**lab demo**)
- Validation using a prototype and simulated context
  - **Parallel group lab experiment:**
    - **Sample of artificial (agents) or natural subjects (students) is constructed/selected**
    - **Effect of treatment with artifact is compared with other treatment**
    - **Statistical discernability of effect in the lab may be established**
    - **Possible explanation in terms of mechanisms**
  - **Lab test**
- Validation using a prototype in a real problem
  - **Parallel group field experiment**
  - **Field test** (technical action research)

Lab test

- Illustration using a small example (not a research method)
- Expert opinion (**opinion survey**)
- Demonstration using a realistic example (**lab demo**)
- Validation using a prototype and simulated context
  - **Parallel group lab experiment**
  - **Lab test**
    - **One case is constructed to contain certain mechanisms (prototype/agents) or selected according to their capabilities (students)**
    - **Realistic context**
    - **Treatment is applied, effects established (possibly using statistics)**
    - **Effects are compared with expectation**
    - **Possible explanation in terms of mechanisms**
- Validation using a prototype in a real problem
  - **Parallel group field experiment**
  - **Field test** (technical action research)

Parallel groups field experiment

- Illustration using a small example (not a research method)
- Expert opinion (**opinion survey**)
- Demonstration using a realistic example (**lab demo**)
- Validation using a prototype and simulated context
  - **Parallel group lab experiment**
  - **Lab test**
- **Validation using a prototype in a real problem**
  - **Parallel group field experiment**
    - **Sample of real subjects is selected in the field (= real context)**
    - **Effect of treatment in the field is compared with other treatment**
    - **Statistical discernability of effect in the field may be established**
    - **Possible explanation in terms of mechanisms**
  - **Field test** (technical action research)

Field test

(a.k.a. technical action research, action case study)

- Illustration using a small example (not a research method)
- Expert opinion (**opinion survey**)
- Demonstration using a realistic example (**lab demo**)
- Validation using a prototype and simulated context
  - **Parallel group lab experiment**
  - **Lab test**
- Validation using a prototype in a real problem
  - **Parallel group field experiment**
  - **Field test** (technical action research, action case study)
    - **The case is a real-world problem**
    - **Researcher uses her technique to solve the problem, or teaches its use to practitioners to solve their problems**
    - **Effects are established (possibly using statistics)**
    - **Effects compared with expectation**
    - **Possible explanation in terms of mechanisms**

- Parallel group experiment: Find differences across treatment groups
- Tests of a sample
- Surveys
- Frequency-based statistical inference all these cases:
  - Assume population distribution & parameters,
  - Predict sample observation ranges,
  - Confirm or reject by actual observations.

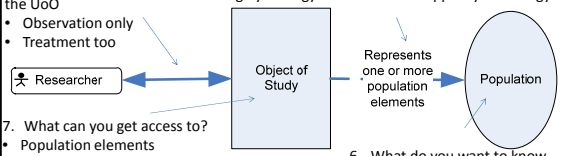
	No treatment by researcher (observational study	Treatment by researcher (experimental study)
Statistical inference	• **Survey**	• **Parallel groups experiments** • **Lab test with sample** • **Field test with sample**
Analogical inference	• **Observational case study (not for validation, but for real-world evaluation)**	• **Lab demo** • **Lab test with case** • **Field test with case**

- Consider architecture of artifact prototype & simulated context
- Observe mechanisms
- Conclude what is likely to happen when implemented artifact is used in practice

Research design decisions

9, 10. How do you want to interact with the UoO
- Observation only
- Treatment too

8. What inference techniques do you want to use?
- Statistical techniques → you need a sample of sufficient size
- Reasoning by analogy → You need to support you analogy



7. What can you get access to?
- Population elements
  - *Real examples of use of BN design techniques*
  - *Real instances of your system*
- Model of population elements (natural or artificial representative)
  - *Real examples of use of other BN design techniques*
  - *Constructed examples of your BN technique*

6. What do you want to know something about?
- *Business network design techniques when used in practice*
- *Your home care system when used in practice*
- *Current physical and digital security techniques*

1. The engineering cycle
2. The technical validation problem
3. Technical validation research questions
4. Validation research methods
5. Examples

Examples from IT theses

- ***DOA estimation for satellite TV reception in cars***
  - ***Artifact:*** *DOA estimation algorithm.*
  - ***Context:*** *arriving satellite TV signals, receiver hardware, beamformer and beamsteering systems, car, driver, passengers.*
  - ***Treatment:*** *Reception of signals, determination of direction while car is moving,* ***in order that*** *passengers can watch satellite TV.*
  - ***Validation questions:***
    - *Effect: Can we recognize DOA?*
    - *Utility: How accurate; space efficient, time efficient?*
    - *Trade-off: Comparison of algorithms?*
    - *Sensitivity to car speed, processor properties?*
  - ***Validation methods:***
    - *Models of DOA algorithms in Matlab, simulated antenna*
    - *Prototype implementations on a Montium 2 processor, simulated antenna*

- ***Information exchange architecture for hospitals and insurance companies***
  - ***Artifact:*** *IT architecture*
  - ***Context:*** *IT of hospitals and insurance companies, business processes, laws and regulations*
  - ***Treatment:*** *Accepting and delivering data both ways,* ***in order that*** *hospitals can get medical activities funded and insurance companies can pay out insurance claims*
  - ***Validation questions:***
    - *Effect: can the required data be exchanged?*
    - *Utility: Does it match currently existing IT architecture in hospitals?*
    - *Trade-offs: No comparison done in the thesis*
    - *Sensitivity analysis: not done in the thesis*
  - ***Validation methods:***
    - *Illustration of architecture on a test case*
    - *Review by experts*

- ***Claims-based workflow management***
  - ***Artifact:*** *claims-based WFM system*
  - ***Context:*** *Employees, business processes, access permissions, business rules*
  - ***Treatment:*** *Process requests to perform activities, respecting access permissions and business rules,* ***in order that*** *employee can perform activities in a flexible manner*
  - ***Validation questions*** *(not asked in the thesis):*
    - *Can claim-based WF handle business processes?*
    - *Utility: can it handle the processes flexibly?*
    - *Trade-off: (Not asked) What overhead is caused by flexibility/*
    - *Sensitivity: (Not asked) Does claim-based WF M scale up to realistic numbers of processes?*
  - ***Validation methods:***
    - *Prototype implementation*
    - *Application to two test cases*

Discussion

- Which validation methods have you used?
- Which validations do you want to perform?

- Wieringa, R.J. (2009) *Design Science as Nested Problem Solving.* In: Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology, Philadelphia. pp. 1-12. ACM. ISBN 978-1-60558-408-9
- Wieringa, R.J. (2010) *Relevance and problem choice in design science.* In: Global Perspectives on Design Science Research (DESRIST). 5th International Conference, 4-5 June, 2010, St. Gallen. pp. 61-76. Lecture Notes in Computer Science 6105. Springer Verlag. ISBN 978-3-642-13334-3