5. Evaluation

Evidence

- In an academic work you need to convince readers that your point of view is correct.
- The only way to do this is to offer credible *evidence*, clearly substantiating the point you are trying to make.
- Evidence is needed whenever you make an assertion or claim that is not self-evidently true to the average reader in your field.
- Evidence is the foundation of any academic argument. Without evidence you don’t have an argument in academic terms – all you have is an opinion.

(Hofstee 2006, pp. 146ff)

Types of Evidence – Appropriate Use of Reseach Method

- Your *research method* determines what type of evidence you will mainly rely on. These may include
  - quotes from literature,
  - statistics from surveys,
  - application data e.g. company financial statements,
  - data from experiments,
  - mathematical calculations,
  - ideas and interpretations from experts,
  - findings from interviews or case studies
  - observations
- Your concrete research usually is made up of different, auxiliary types of evidence
- Appropriate use of the research method provides evidence

(Hofstee 2006, pp. 146ff)

Evidence is based on Data and Analysis

- Your research *provides you with relevant facts or data* that you can *analyse and use as evidence* to prove your thesis statement (resp. answer your research question)
- If you want your readers to accept or even consider your argument, you need
  - the **data** to substantiate your point and
  - provide **analysis** and argumentation that gives meaning to the data

(Hofstee 2006, pp. 146ff)

Evidence depends on the quality of data and the use of it

- All data and facts can be used as evidence for something, and they can be used effectively or ineffectively.
- After you have established what type of evidence is appropriate to your dissertation, it is up to you to make sure that the evidence you present is both
  - of sufficient quality and
  - used appropriately

(Hofstee 2006, pp. 146ff)

Sufficient Data Quality

- «Of sufficient quality» means that …
  - … the data is realiable,
  - … there is enough of it,
  - … it pertains directly to your point, and
  - … it is current

    (i.e. it must not have been superseded by later evidence that has established the earlier work to be inapplicable or flawed)

(Hofstee 2006, pp. 146ff)

The quality of your data will determine the quality of your dissertation ..

Using Evidence Appropriately

- Your use of the evidence must
  - relate it clearly to the point that you want to make.
- If your evidence has shortcomings, but you still believe it is useful, you must
  - admit to those shortcomings and
  - justify why you believe the evidence still supports your point.
- You should present both
  - evidence that is in favor of your thesis and
  - evidence that contradicts it.

  (Ignoring contradicting evidence or weaknesses in evidence suggests to readers that you are afraid that your argument would fall apart.)

(Hofstee 2006, pp. 146ff)

Evidence in particular Research Methods

Evidence and Research Methods

- Research methods differ in how they gather and use data.
- To provide evidence you have to make sure that you
  - the research method was suitable for your research
  - have applied it appropriately and
  - that the conclusions you drew are valid
- The reader and reviewer of your thesis/paper (in particular your supervisor) will check whether you did it adequately
- In the following we provide some sample questions you (or a reviewer) can use as guidelines to check the adequate application of the research method.

Sample Questions for Evaluating Applications of Research Methods(1)

Survey (Quantitative)

- Was the sample size big enough?
- What information is given about the response rate?
- Did the researchers make efforts to see if there were significant differences between respondents and non-respondents?
- Do the researchers use the survey results to make generalizations about larger population? Is this appropriate?
- What limitations in their survey strategy do the researchers recognize?

Questionnaire (Qualitative)

- Are the questions appropriate to get data for the research?
- What question types where used – open, closed, both? Was this appropriate?
- Are the questions and possible responses clear, unambiguous, the appropriate format, in the right order?
- Do the researchers discuss content validity, construct validity and reliability of their questionnaire? If not, how does this affect the confidence in the research?

Selected from (Oates 2010)

Sample Questions for Evaluating Applications of Research Methods (2)

Interviews

- What information is given about the interviewer and how they might have affected the interview? Is this sufficient?
- Are sufficient quotations from the interviews used in the report?
- Do the researchers use the interview findings to make generalizations about larger population? Is this appropriate?
- What limitations in their interviews do the researchers recognize?

Observations

- Where the items observed easily observable, unambiguous and independent from each other?
- Did they occur regularly enough to provide sufficient data?
- Were the items observed the most appropriate for the research objectives.
- Was the time spent in the field long enough?
- Did the research avoid disrupting the naturalness of the setting?
- What limitations in the use of observation does the researcher recognize?

Selected from (Oates 2010)

Sample Questions for Evaluating Applications of Research Methods (3)

Case Study

- Have the criteria for choosing the particular case(s) been described and justified?
- Was the time spent in the field long enough?
- Does the research look at relationships and processes and provide a holistic perspective?
- What kind of generalizations are reported, if any?
- Are the generalizations appropriate?

Design Research

- What information is given about the development methodology? Is the methodology appropriate?
- Do the researchers discuss all stages of the systems development life-cycle or just some stages?
- **What do the researchers tell about how they evaluated their artifact?**
- **What evaluation criteria do they use? Are these appropriate?**
- Do the researchers use their results to make generalizations about the use of their artifact in other situations? Is this appropriate?

Selected from (Oates 2010)

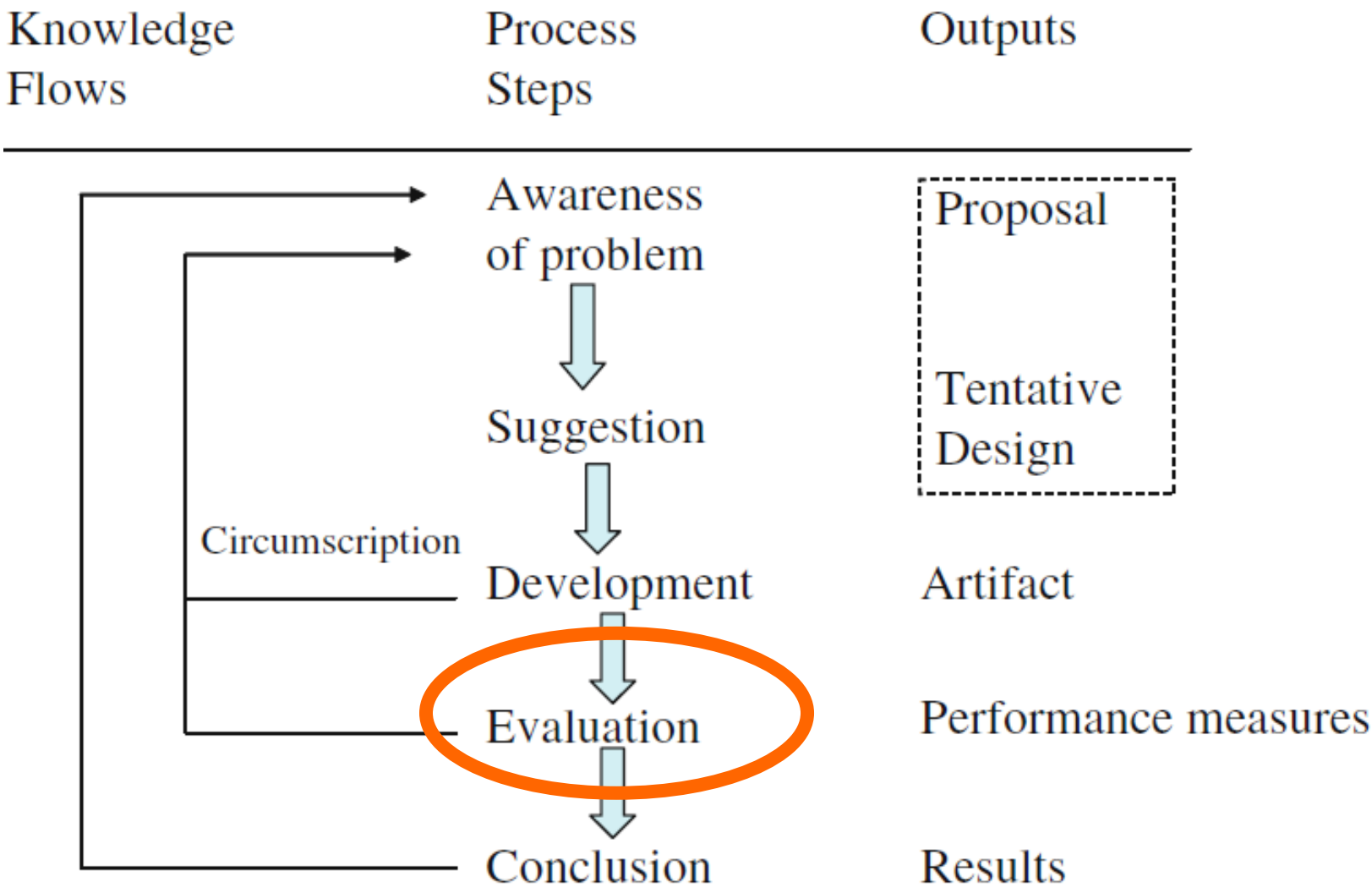
Evaluation: Evidende for Design Research

Evaluation in Design Science Research

- Evidence in design research is specific.
- Based on the collected data, a design researcher designs an artifact that provides utility
- In addition the researcher has to provide evidence that this artifact solves a real problem.
- To provide evidence, it has to be examined whether the artifact meets the requirements. This is done in the **evaluation phase**.
- Evidence-based artifact evaluation requires that the artifact is evaluated within the business environment.
- A design science paper with no evaluation is least likely to be accepted for a conference or journal

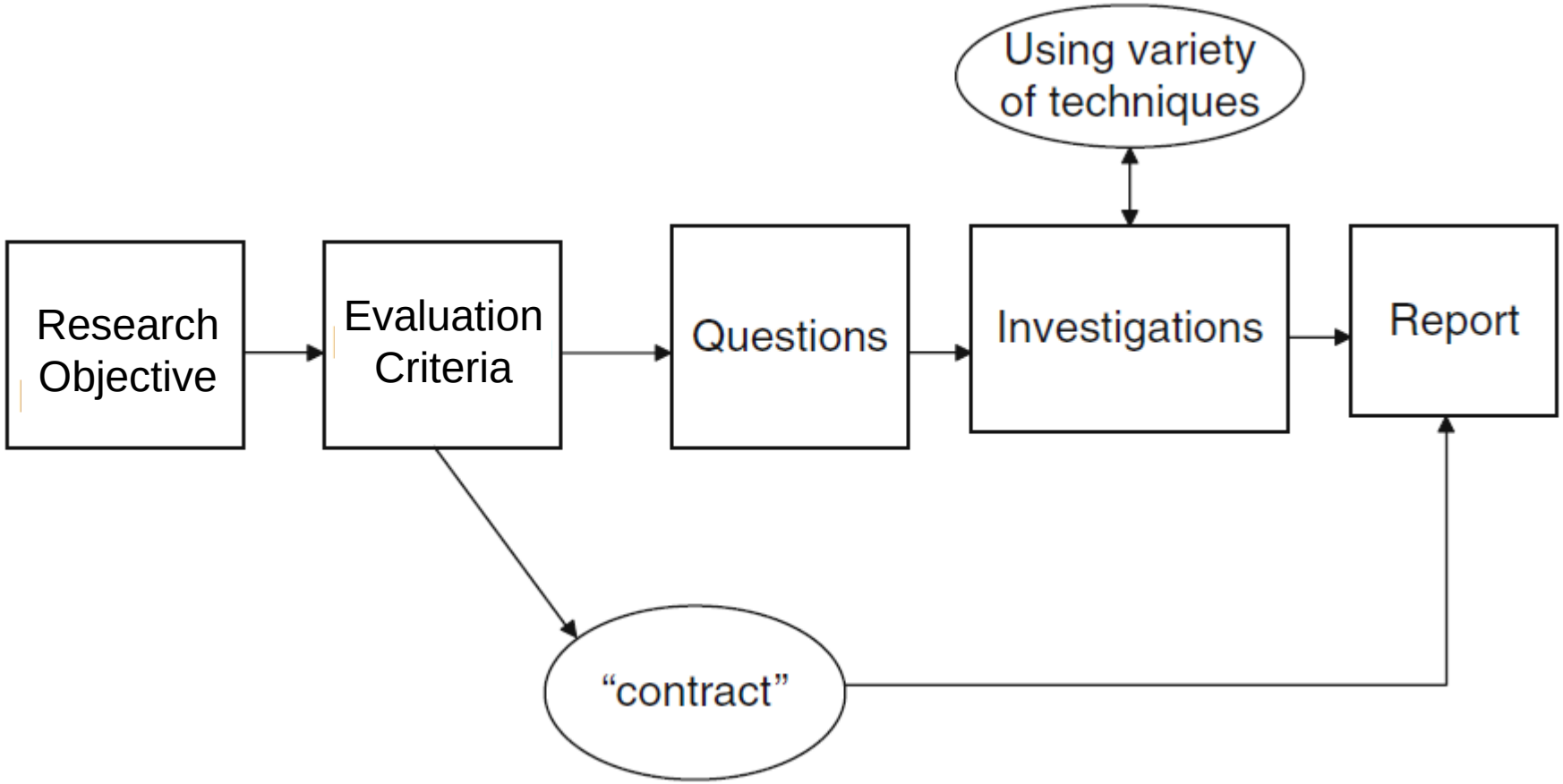
(Hevner & Chatterjee 2010, p. 110/122)

Design Science Research Framework



(Hevner & Chatterjee 2010, p. 27)

Structure of an Evaluation Study



Evaluation should be viewed as an exercise in argument, rather than as a demonstration, because any study appears equivocal when subjected to serious scrutiny.

adapted from (Hevner & Chatterjee 2010, p. 112f)

What to evaluate in Design Science Research

- Depending on the artifact, there are different aspects that can be evaluated, for example
  - ♦ performance (of a technical system)
  - ♦ organizational impact
- Depending on what to be evaluated there are different evaluation methods, e.g.

What to evaluate	Evaluation Methods
Performance	analytical modeling, simulation, measurement, testing
Organisation impact	quantitative surveys, qualitative interviews, focus groups, questionnaires, observation

How to do evaluation?

Example1: *We develop and introduce a new IT system that supports customer consultants in a bank in recommending financial products reducing effort and making recommendations better*

- How would you make the evaluation?

Evaluation Methods (1)

- *Measurements / Observational Case Studies*
  - study the designed artifact in depth in a real environment.
  - observe criteria and monitor the use of the artifact to gain understanding of its value and utility.
  - Measurement is typically used for Performance Evaluation
    - Comparing your system with other similar systems
    - Before-after comparison
    - Determining the optimal value of a parameter (system tuning, workload)
    - Predicting the performance at future loads (scaling and forecasting).
  - *Metrics* are criteria for dependent and independent variables to evaluate the performance of the system, e.g.
    - efficiency (time required, use of resources, scalability)
    - effectiveness (accuracy, quality of results)

Problems with Measurements

- Data not available
  - To analyse the effect of an artefact it often has to be compared to previous situation for which historic data is missing
- Artefact is not executable (framework, concept)

Additional problems for measuring organisational impact :

- Putting system into practice is not possible
  - Changing processes or organisation structure required
  - Technical integration of a system is additional effort
- Not enough time to make the measurements
  - Observation might require long durations of months or weeks to observe the effect of the artefact

Evaluation Methods (2)

Alternative evaluations, if you cannot measure the artefact in a real environment:

- *Descriptive* evaluation
  - Informed argument uses information from knowledge base to build a convincing argument for artifact's utility.
- *Scenarios construction*
  - construct detailed scenarios around artifact to demonstrate its utility.
- *Experimental* methods
  - controlled experiments in which you study the artifact in controlled environment for qualities (e.g., usability).
  - *simulation* models: execute the artifact with artificial data and observe dynamic performance behavior and scalability.
- *Analytical* techniques
  - examine the structure of the artifact for static qualities (e.g., complexity, architecture) or behavious

Discussion

- In a research thesis a student developed a framework for building up a lessons learned knowledge base
- The framework consisted of a description of the necessary aspects of lessons learned systems, a procedure model specifying the steps to develop such a system and a set of recommendations for each step
- In order to evaluate the framework, the student used a qualitative approach
  - He sent the description of the framework to several practitioners
  - He made interviews with the practitioners using a questionnaire
    - to find out, whether the framework is comprehensible
    - to get an assessment whether the framework would help to develop useful knowledge bases
- Discuss the appropriateness of the evaluation
  - Why did the student choose this evaluation?
  - What are the weaknesses?

Critical Evaluation

- The evaluation should be critical in order to provide evidence
- Even if not applicable in a real scenario, you have to allow and stimulate a critical assessment of your artefact
  - Construction a scenario in which the systems is used and assessed
  - Make a workshop in which the artefact is evaluated be several people (maybe covering different perspective), e.g. focus group