Evaluating the Outcome of Modeling Research

Prof. Dr. Barbara Weber
About Myself

2000-2016

universität innsbruck

2016-2019

DTU

since 2019

University of St. Gallen
Agenda

• Modeling Research as Design Science Research
• Types of Design Artifacts
• Three Cycle View of Design Science Research
• Methods for Evaluating Design Science Research
• Evaluation Framework for Design Science Research
Modeling research deals with all aspects of modeling, from languages and methods, to tools and applications.
Modeling research deals with all aspects of modeling, from languages and methods, to tools and applications.

Artifacts play a crucial role in modeling research.
Modeling Research as Design Science Research (DSR)

Modeling research deals with all aspects of modeling, from languages and methods, to tools and applications.

Artifacts play a crucial role in modeling research.

Design science research focuses on the development and performance of (designed) artifacts with the explicit intention of improving the functional performance of the artifact.
## Types of Design Artifacts

<table>
<thead>
<tr>
<th>Artifact Type</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct</td>
<td>Vocabulary and symbols to define and understand problems and solutions; constitute the language to specify problems and solutions</td>
<td>Routing symbols (and other modeling primitives) in the process modeling domain (Figl et al. 2013)</td>
</tr>
</tbody>
</table>

Source: Hevner et al. 2004
### Example: Routing Symbol Design

<table>
<thead>
<tr>
<th>AND</th>
<th>R_{UML}</th>
<th>R_{BPMN}</th>
<th>R_{EPC}</th>
<th>R_{YAWL}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer Shape</td>
<td></td>
<td>symmetric diamond-shape</td>
<td>circle</td>
<td>rectangle</td>
</tr>
<tr>
<td>Inner Shape</td>
<td>narrow rectangle (bar)</td>
<td></td>
<td>logical marker for 'and' (\textasciicircum)</td>
<td>left- and right-sided open triangle</td>
</tr>
<tr>
<td>XOR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer Shape</td>
<td>diamond-shape without internal marker</td>
<td>symmetric diamond-shape</td>
<td>circle</td>
<td>rectangle</td>
</tr>
<tr>
<td>Inner Shape</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Routing symbols of four existing process modeling languages

Source: Figl et al. 2013
# Types of Design Artifacts

<table>
<thead>
<tr>
<th>Artifact Type</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construct</strong></td>
<td>Vocabulary and symbols to define and understand problems and solutions; constitute the language to specify problems and solutions</td>
<td>Routing symbols (and other modeling primitives) in the process modeling domain (Figl et al. 2013)</td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td>Designed representations of problem and/or solution;</td>
<td>Process models represented in BPMN; ER models; Conceptual framework to support automated product configuration in cyber physical systems (Safdar et al. 2020)</td>
</tr>
</tbody>
</table>

Source: Hevner et al. 2004
Example: Framework for Product Configuration in Cyber Physical Systems (CPS)

Conceptual framework to support multi-stage and multi-step automated product configuration of CPSs. The framework includes classification of constraints and a list of automated functionalities of a CPS configuration solution.

Source: Safdar, S.A., Lu, H., Yue, T. et al. 2021
### Types of Design Artifacts

<table>
<thead>
<tr>
<th>Artifact Type</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct</td>
<td>Vocabulary and symbols to define and understand problems and solutions; constitute the language to specify problems and solutions</td>
<td>Routing symbols (and other modeling primitives) in the process modeling domain (Figl et al. 2013)</td>
</tr>
<tr>
<td>Model</td>
<td>Designed representations of problem and/or solution</td>
<td>Process models represented in BPMN; ER models; Conceptual framework to support automated product configuration in cyber physical systems (Safdar et al. 2020)</td>
</tr>
<tr>
<td>Method</td>
<td>Algorithms, practices, and recipes for performing a task</td>
<td>Test-driven modeling (Zugal et al. 2013); Structured Process Modeling Method (Claes et al. 2017)</td>
</tr>
</tbody>
</table>

Source: Hevner et al. 2004
Example: Test-driven Modeling

To overcome problems in understanding and maintaining declarative process models Test-driven Modeling (a novel modeling method) has been proposed.

An implementation of the concepts of TDM are provided by Test Driven Modeling Suite.

Source: Zugal et al. 2013
Example: Structured Process Modeling Method

Conceptual method to derive a modeler's cognitive profile and the related optimal modeling strategy.

A system for operationally supporting the method through automated modeling strategy selection and training.

Source: Claes et al. 2017
## Types of Design Artifacts

<table>
<thead>
<tr>
<th>Artifact Type</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construct</strong></td>
<td>Vocabulary and symbols to define and understand problems and solutions; constitute the language to specify problems and solutions</td>
<td>Routing symbols (and other modeling primitives) in the process modeling domain (Figl et al. 2013);</td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td>Designed representations of problem and/or solution</td>
<td>Process models represented in BPMN; ER models; Conceptual framework to support automated product configuration in cyber physical systems (Safdar et al. 2020)</td>
</tr>
<tr>
<td><strong>Method</strong></td>
<td>Algorithms, practices, and recipes for performing a task</td>
<td>Test-driven modeling (Zugal et al. 2013); Structured Process Modeling Method (Claes et al. 2017)</td>
</tr>
<tr>
<td><strong>Instantiation</strong></td>
<td>Physical realizations that act on the natural world (implemented systems, prototypes)</td>
<td>Prototype for runtime flexibility during data-centric and data-driven process execution (Andrews et al. 2019)</td>
</tr>
</tbody>
</table>
Example: Run-time flexibility for data-aware and data-driven processes

Concepts and algorithms for supporting ad-hoc changes during data-centric and data-driven process execution.

A proof-of-concept implementation and its application to various applications.

Source: Andrews et al. 2019
Design Science Research
Three Cycle View

Environment

Application Domain
- People
- Organizational Systems
- Technical Systems
- Problems & Opportunities

Relevance Cycle
- Requirements
- Field Testing

Design Science Research

Build Design Artifacts & Processes
Design Cycle
Evaluate

Knowledge Base (KB)

Foundations
- Scientific Theories & Methods
- Experience and Expertise
- Meta-Artifacts (Design Products & Design Processes)

Source: Hevner et al. 2007
Design Science Research
Three Cycle View

Environment

Application Domain
- People
- Organizational Systems
- Technical Systems
- Problems & Opportunities

Relevance Cycle
- Requirements
- Field Testing

Design Science Research

Build Design Artifacts & Processes

Design Cycle

Evaluate

Knowledge Base (KB)

Foundations
- Scientific Theories & Methods
- Experience and Expertise
- Meta-Artifacts (Design Products & Design Processes)

Source: Hevner et al. 2007
Methods for Evaluating Design Science Research (DSR)

Conceptual Evaluation

discusses the artifact’s strength, weaknesses, and limitations

Empirical Evaluation

empirically evaluates the artifact using, for example, controlled experiments, case studies, action research, quantitative simulation, a benchmarking study

Source: https://github.com/acmsigsoft/EmpiricalStandards/blob/master/docs/EngineeringResearch.md
Methods for Evaluating Design Science Research (DSR)

Conceptual Evaluation

discusses the artifact’s strength, weaknesses, and limitations

Empirical Evaluation

empirically evaluates the artifact using, for example, controlled experiments, case studies, action research, quantitative simulation, a benchmarking study

Source: https://github.com/acmsigsoft/EmpiricalStandards/blob/master/docs/EngineeringResearch.md
Example: Conceptual Evaluation of Process Modeling Languages

Source: Recker et al. 2007
Completeness of a description can be measured as:

- Construct deficit
- Construct overload
- Construct redundancy
- Construct excess

Source: Recker et al. 2007
Example: Conceptual Evaluation of Process Modeling Languages

Patterns

- Workflow patterns (Russell et al. 2016)
- Change patterns (Weber et al. 2008)
- Time patterns (Lanz et al. 2014)
- VIVACE framework (Ayora et al. 2015)

Source: Recker et al. 2007
Example: Conceptual Evaluation of Modeling Languages

Overview of Identified Time-Patterns

Time patterns aim at the comparison of technologies for realizing time-and process-aware information systems. Moreover, they aim to provide a reference for implementing time support.

Source: Lanz et al. 2014
Example: Conceptual Evaluation of Modeling Languages

Overview of Identified Time-Patterns

Time patterns aim at the comparison of technologies for realizing time-and process-aware information systems. Moreover, they aim to provide a reference for implementing time support.

Source: Lanz et al. 2014
Example: Framework for Product Configuration in Cyber Physical Systems (CPS)

Conceptual framework to support multi-stage and multi-step automated product configuration of CPSs

Framework can serve as guide to researchers and practitioners to evaluate an existing CPS-specific PLE solution or devise a new one.

Source: Safdar, S.A., Lu, H., Yue, T. et al. 2021
Methods for Evaluating Design Science Research (DSR)

Conceptual Evaluation

discusses the artifact’s strength, weaknesses, and limitations

Empirical Evaluation

depicts the artifact using, for example, controlled experiments, case studies, action research, quantitative simulation, a benchmarking study

Source: https://github.com/acmsigsoft/EmpiricalStandards/blob/master/docs/EngineeringResearch.md
Empirical Methods for Evaluating DSR

- **Controlled Experiment**: human participants use the artifact
- **Case Study**: researchers observe a real organization using the artifacts
- **Action Research**: researchers intervene in a real organization using the artifact
- **Quantitative Simulation**: artifact is assessed (usually against a competing artifact) in an artificial environment
- **Benchmarking Study**: artifact is assessed using one or more benchmarks

Source: https://github.com/acmsigsoft/EmpiricalStandards/blob/master/docs/EngineeringResearch.md
# Variables for the Evaluation of DSR Artifacts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Approach</strong></td>
<td>Qualitative</td>
</tr>
<tr>
<td><strong>Artifact Focus</strong></td>
<td>Technical, Organizational, Strategic</td>
</tr>
<tr>
<td><strong>Artifact Type</strong></td>
<td>Construct, Model, Method, Instantiation, Theory</td>
</tr>
<tr>
<td><strong>Epistemology</strong></td>
<td>Positivism, Interpretivism</td>
</tr>
<tr>
<td><strong>Function</strong></td>
<td>Knowledge function, Control function, Development function, Legitimization function</td>
</tr>
<tr>
<td><strong>Method</strong></td>
<td>Action research, Case study, Field experiment, Formal proofs</td>
</tr>
<tr>
<td></td>
<td>Controlled experiment, Prototype, Survey</td>
</tr>
<tr>
<td><strong>Object</strong></td>
<td>Artifact, Artifact construction</td>
</tr>
<tr>
<td><strong>Ontology</strong></td>
<td>Realism, Nominalism</td>
</tr>
<tr>
<td><strong>Perspective</strong></td>
<td>Economic, Deployment, Engineering, Epistemological</td>
</tr>
<tr>
<td><strong>Position</strong></td>
<td>Externally, Internally</td>
</tr>
<tr>
<td><strong>Reference Point</strong></td>
<td>Artifact against research gap, Artifact against real world, Research gap against real world</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>Ex ante, Ex post</td>
</tr>
</tbody>
</table>

Source: Cleven et al. 2009
Evaluating Design Science Research
Two Distinct Goals

Goal Type I

Demonstrate that a very new artifact works, i.e., a solution to an unsolved problem was found.

Goal Type II

Demonstrate that the design artifact works better than existing solutions, e.g., the artifact solves a problem much more efficiently and with fewer resources.

Source: Mettler et al. 2014
Example: Framework for Product Configuration in Cyber Physical Systems (CPS)

Assess if the framework provides the support for capturing and managing commonalities, variabilities, and constraints in the domain engineering phase as well as the support for automation of configuration in the application engineering phase.

Source: Safdar, S.A., Lu, H., Yue, T. et al. 2021
Example: Structured Process Modeling Method

Assess if selecting and training modelers in their optimal process modeling strategy (i.e., the proposed Structured Modeling Method) improves modeling performance.

Source: Claes et al. 2017
Example: Test-driven Modeling

Assess the impact of Test-driven Modeling on communication behavior (communication between domain experts and model builders) using a case study.

Assess the impact of Test-driven Modeling on maintainability using a controlled experiment.

Source: Zugal et al. 2013
Example: Run-time flexibility for data-aware and data-driven processes

Assess the **feasibility** of ad-hoc changes without disruptions.

Show that changes are **possible** concerning every aspect of a process model.

Evaluate the **scalability** of the approach.

Source: Andrews et al. 2019
Example: Routing Symbol Design

<table>
<thead>
<tr>
<th></th>
<th>$R_{UML}$</th>
<th>$R_{BPMN}$</th>
<th>$R_{EPC}$</th>
<th>$R_{YAWL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AND</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer Shape</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inner Shape</td>
<td>narrow rectangle (bar)</td>
<td>symmetric diamond-shape</td>
<td>logical marker for 'and' ((^{\lor}))</td>
<td>left- and right-sided open triangle</td>
</tr>
<tr>
<td><strong>XOR</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outer Shape</td>
<td>diamond-shape without internal marker</td>
<td>symmetric diamond-shape</td>
<td>circle</td>
<td>rectangle</td>
</tr>
<tr>
<td>Inner Shape</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparing the effect of different routing symbol designs of four different pre-existing modeling languages in the context of process model comprehension tasks.

Source: Figl et al. 2013
Empirical Methods for Evaluating DSR

**Controlled Experiment**
human participants use the artifact

**Case Study**
researchers observe a real organization using the artifacts

**Action Research**
researchers intervene in a real organization using the artifact

**Quantitative Simulation**
artifact is assessed (usually against a competing artifact) in an artificial environment

**Benchmarking Study**
artifact is assessed using one or more benchmarks

Use of Experiments in Design Science Research

Source: Mettler et al. 2014

1. Improve design product (artifact)
2. Improve design process (build/test cycle)
3. Develop design theory
4. Stop (build/test cycle)
Use of Experiments in Design Science Research

Definitions:
- Randomization: Defines randomization process and scope.
- Manipulation: Defines manipulation procedure and metrics.
- Data Analysis: Defines data analysis procedure.

Experimental Setting:
- Study Subjects: test group, control group.
- Constructs, Models, Methods, Instantiations: alternative 1, ..., alternative n.
- Experimental Setting: are being manipulated on.

Experimental Results:
- Randomly assigns subjects to groups.
- Examines, observes.
- Calculates, interprets.

Source: Mettler et al. 2014

Objectives:
1. Improve design product (artifact).
2. Improve design process (build/test cycle).
3. Develop design theory.
4. Stop (build/test cycle).
Experiment Principles

Theory

- Cause Construct
- Effect Construct

Experiment Objective

- Cause
- Effect

Observation

- Treatment
- Output

Experiment Operation

- Independent Variable
- Dependent Variable

Wohlin et al. 2012
Example: Routing Symbol Design

Effect of routing symbol design (i.e., perceptual discriminability, pop out, semantic transparency, and aesthetic) design during model comprehension task on model comprehension accuracy, efficiency, and perceived cognitive load.

Source: Figl et al. 2013
### Example: Routing Symbol Design

<table>
<thead>
<tr>
<th>AND</th>
<th>R_{shape}</th>
<th>R_{sym}</th>
<th>R_{per}</th>
<th>R_{disc}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer Shape</td>
<td>narrow rectangle (bar)</td>
<td>symmetric diamond-shape</td>
<td>circle</td>
<td>rectangle</td>
</tr>
<tr>
<td>Inner Shape</td>
<td>-</td>
<td>internal marker (“•”)</td>
<td>logical marker for ‘and’ (‘^’)</td>
<td>left- and right-sided open triangle</td>
</tr>
<tr>
<td>XOR</td>
<td>diamond-shape without internal marker</td>
<td>symmetric diamond-shape</td>
<td>circle</td>
<td>rectangle</td>
</tr>
</tbody>
</table>

Effect of **routing symbol design** (i.e., perceptual discriminability, pop out, semantic transparency, and aesthetic) design during model comprehension task on model comprehension accuracy, efficiency, and perceived cognitive load.

**Results:** Design principles related to
- Perceptual discriminability and pop out improve comprehension accuracy
- Semantic transparency and aesthetic design of symbols lower perceived difficulty of comprehension

Source: Figl et al. 2013
Example: Test-driven Modeling

Effect of Test-driven modeling on perceived cognitive load, perceived quality, and quality of the adapted models in the context of maintainability tasks.

Source: Zugal et al. 2013
Example: Test-driven Modeling

Effect of Test-driven modeling on perceived cognitive load, perceived quality, and quality of the adapted models in the context of maintainability tasks.

Results: The adoption of test cases
- could significantly reduce the perceived cognitive load
- could significantly improve the perceived quality
- while no significant effects on model quality could be shown

Source: Zugal et al. 2013
Utility emerges through the use of the artifact and depends on the user and the environment.

Utility is relative!

Source: Mettler et al. 2014
Evaluation Framework for DSR

1. User

Study Subjects
Experimental Setting

Source: Mettler et al. 2014
Evaluation Framework for DSR: User

- **User** as the centerpiece of the evaluation framework
- The user determines **how** an artifact is **used** and what **value** she can gain from it

**Example:** *Experiment to investigate the impact of two distinct business process modeling notations on model comprehension.*
Evaluation Framework for DSR: User

- **User** as the centerpiece of the evaluation framework
- The user determines **how** an artifact is **used** and what **value** she can gain from it

**Example:** Experiment to investigate the impact of **two** distinct business process modeling notations on model comprehension.
Evaluation Framework for DSR: User

- **User** as the centerpiece of the evaluation framework
- The user determines **how** an artifact is **used** and what **value** she can gain from it

**Example:** Experiment to investigate the impact of two distinct business process modeling notations on model comprehension.

**Study Subjects and Experimental Setting:**
- Are study subject similarly knowledgeable in both modeling notations?
Example: Imperative vs. Declarative Process Models

Controlled experiment comparing two modeling notations and two different task types in terms of accuracy and modeling speed.

Example of an imperative model

Example of a declarative model

Rather familiar  
Limited familiarity

Source: Pichler et al. 2011
Evaluation Framework for DSR: User

• **User** as the centerpiece of the evaluation framework
• The user determines **how** an artifact is **used** and what **value** she can gain from it

*Example: Experiment to investigate the impact of a particular method on modeling performance of master students.*
Evaluation Framework for DSR: User

• **User** as the centerpiece of the evaluation framework
• The user determines **how** an artifact is **used** and what **value** she can gain from it

**Example:** Experiment to investigate the impact of a particular method on modeling performance of **master students**.

**Study Subjects and Experimental Setting:**
• Are study participants representative for the group of user for which the artifact was developed?
Example: Structured Process Modeling Method

Method developed to support modelers.

146 master students of the Business Engineering program at Ghent University (Belgium) participated in the experiment.

Paper justifies why master students have been chosen (instead of modeling practitioners or younger students).

Source: Claes et al. 2017
Evaluation Framework for DSR

Goals and scope of usage
Artifact characteristics
Manipulation procedure

Study Subjects
Experimental Setting

Source: Mettler et al. 2014
The use of the artifact is another crucial piece of information, since the situation on how the artifact is used influences its utility.

Goal Type I
- Demonstrate that totally new artifact works

Goal Type II
- Demonstrate that developed artifact is better than existing ones
Evaluation Framework for DSR: Use

Example: “A study in the area of business intelligence (BI) identifies a major problem in the representation of data. A prototype was developed in order to provide new means for visualizing data. Part-time MBA students were asked to perform distinct predefined tasks with the aim of comparing a traditional visualization with the newly developed representation. A questionnaire was used to capture the participants’ personal beliefs on the usability of the solution.”

Source: Mettler et al. 2014
Evaluation Framework for DSR: Use

**Example:** “A study in the area of business intelligence (BI) identifies a major problem in the representation of data. A prototype was developed in order to provide new means for visualizing data. Part-time MBA students were asked to perform distinct predefined tasks with the aim of comparing a traditional visualization with the newly developed representation. A questionnaire was used to capture the participants’ personal beliefs on the usability of the solution.”

**Scope and design goals are often reported in insufficient detail:**
- For which type of data, which type of user, which domain, and which design goal (e.g., cost, quality, efficiency) was the artifact developed?
Evaluation Framework for DSR: Use

Example: “With the aim of improving the learning process of software developers, a novel method was designed which integrates additional information into an existing programming environment. The utility of the artifact is measured by a couple of metrics, such as the number of correctly answered questions related to a defined problem, the amount of time for finding deficiencies in programming code, or the total number of found deficiencies. The necessary data for evaluating the method was obtained from a questionnaire consisting of multiple-choice questions related to practical programming problems which was answered by undergraduate students.”

Source: Mettler et al. 2014
Evaluation Framework for DSR: Use

**Example**: “With the aim of improving the learning process of software developers, a novel method was designed which integrates additional information into an existing programming environment. The utility of the artifact is measured by a couple of metrics, such as the number of correctly answered questions related to a defined problem, the amount of time for finding deficiencies in programming code, or the total number of found deficiencies. The necessary data for evaluating the method was obtained from a questionnaire consisting of multiple-choice questions related to practical programming problems which was answered by undergraduate students.”

**Artifact characteristics and manipulation procedure:**

- Details concerning artifact characteristics and manipulation procedure is missing (who does what, when, where, and how)
- What programming problems were asked? How were the questions asked? In which situations were the students allowed to use the new method? What other auxiliary materials did the students have?

Source: Mettler et al. 2014
## Evaluation Framework for DSR: Use

### Overview of empirical studies into hierarchical structuring

<table>
<thead>
<tr>
<th>Work</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moody [15]</td>
<td>Positive influence on accuracy, no influence / negative influence on time</td>
</tr>
<tr>
<td>Domain: ER-Models</td>
<td></td>
</tr>
<tr>
<td>Reijers et al. [16][17]</td>
<td>Positive influence on understandability for one out of two models</td>
</tr>
<tr>
<td>Domain: Business Process Models</td>
<td></td>
</tr>
<tr>
<td>Cruz-Lemus et al. [9][18]</td>
<td>Series of experiments, positive influence on understandability in last experiment</td>
</tr>
<tr>
<td>Domain: UML Statecharts</td>
<td></td>
</tr>
<tr>
<td>Cruz-Lemus et al. [13]</td>
<td>Hierarchy depth of statecharts has no influence</td>
</tr>
<tr>
<td>Domain: UML Statecharts</td>
<td></td>
</tr>
<tr>
<td>Shoval et al. [14]</td>
<td>Hierarchy has no influence</td>
</tr>
<tr>
<td>Domain: ER-Models</td>
<td></td>
</tr>
<tr>
<td>Cruz-Lemus et al. [8]</td>
<td>Positive influence on understandability for first experiment, negative influence in replication</td>
</tr>
<tr>
<td>Domain: UML Statecharts</td>
<td></td>
</tr>
<tr>
<td>Cruz-Lemus et al. [12][19]</td>
<td>Hierarchy depth has a negative influence</td>
</tr>
<tr>
<td>Domain: UML Statecharts</td>
<td></td>
</tr>
</tbody>
</table>

Source: Zugal et al. 2011
## Evaluation Framework for DSR: Use

### Overview of Empirical studies into hierarchical structuring

<table>
<thead>
<tr>
<th>Work</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moody [15]</td>
<td>Positive influence on accuracy, no influence / negative influence on time</td>
</tr>
<tr>
<td>Reijers et al. [20]</td>
<td>Reliability for one out of two structures</td>
</tr>
<tr>
<td>Cruz-Lemus [9]</td>
<td>Stability of the influence on understanding</td>
</tr>
<tr>
<td>Cruz-Lemus [12]</td>
<td>Stability of the influence on accuracy, negative influence for first replication</td>
</tr>
<tr>
<td>Shoval et al. [19]</td>
<td>Stability of the influence for first replication</td>
</tr>
<tr>
<td>Cruz-Lemus et al. [12][19]</td>
<td>Artifact use has a negative influence, negative influence in replication</td>
</tr>
</tbody>
</table>

Source: Zugal et al. 2011
The Importance of the Task

Source: Zugal et al. 2011
The Importance of the Task

- Question complexity induces abstraction, which lowers mental effort. Mental effort determines performance.
- Abstraction enables hierarchy, which causes split-attention effect.

Source: Zugal et al. 2011
The Importance of the Task

Are activities J and K mutually exclusive?

Source: Zugal 2013
The Importance of the Task

Are activities B and F mutually exclusive?

Source: Zugal 2013
Key lesson: Results depend on *which* questions are asked or on a more abstract note on their *use*.

Source: Zugal et al. 2011, Zugal 2013
**Evaluation Framework for DSR: Use**

**Key lesson:** Results depend on *which* questions are asked or on a more abstract note on their *use*.

- Precise description of **artifact characteristics** and **manipulation procedure** needed for replication. Ideally, a replication package can be provided.
Evaluation Framework for DSR

1. User
2. Use
3. Utility

Evaluation metrics
Evaluation results

Goals and scope of usage
Artifact characteristics
Manipulation procedure

Study Subjects
Experimental Setting

Source: Mettler et al. 2014
Evaluation Framework for DSR: Utility

- **Utility** emerges through the use of the artifact and depends on the user and the environment
- **Utility is a relative concept**
- Clear and measurable variables are needed to assess utility
- Chosen metrics should give alternatives that are compared equal consideration (different alternatives might have been designed with different goals in mind)
Evaluation Framework for DSR: Utility

• Keep moderating and mediating effects in mind
  • User-specific characteristics (e.g., age, gender, and computer literacy)
  • Date and time (e.g., differences in bandwidth utilization depending on specific workdays)
  • Technical effects (e.g., divergent behavior of the designed artifact on different platforms)
  • Environmental effects (e.g., divergent behavior of the designed artifact due to temperature differences)
  • Socio-cultural effects (e.g., assignment of distinct connotations and meaning for the same artifact construct because of a different cultural background)
Evaluation Framework for DSR: Utility

Example: “A study describes a new search algorithm for maximizing the proportion of useful hits. A design experiment was conducted with the aim to proof that the new algorithm provides more useful results than the hits of a commercial search engine. The “utility” was judged by means of user feedback. The metrics to measure search performance are “elapsed time for presenting search result” and “selectivity of responses,” Metrics to describe the search quality are “number of good sources” (as defined by the user), “number of duplicates in results list,” and “average list length.”

Source: Mettler et al. 2014
Evaluation Framework for DSR: Utility

Example: “A study describes a new search algorithm for maximizing the proportion of useful hits. A design experiment was conducted with the aim to proof that the new algorithm provides more useful results than the hits of a commercial search engine. The “utility” was judged by means of user feedback. The metrics to measure search performance are “elapsed time for presenting search result” and “selectivity of responses.” Metrics to describe the search quality are “number of good sources” (as defined by the user), “number of duplicates in results list”, and “average list length”.

- Importance of mediating and moderating factors:
  - Mediating and moderating factors play an important role in the scenario above (e.g., goodness of hits might be assessed differently by users depending on contextual and situational factors).

Source: Mettler et al. 2014
Evaluation Framework for DSR

Utility emerges through the use of the artifact and depends on the user and the environment.

Utility is relative!

Source: Mettler et al. 2014
Excellent Resources Available

- Design Science Methodology
- Experimentation in Software Engineering
References

- Anne Cleven, Philipp Gubler, Kai M. Hüner: 
  *Design alternatives for the evaluation of design science research artifacts.* DESRIST 2009

- Alan R. Hevner, Salvatore T. March, Jinsoo Park, Sudha Ram: 

- Alan R. Hevner: 

- Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell: 

- Kathrin Figl, Jan Recker, Jan Mendling: 
  *A study on the effects of routing symbol design on process model comprehension.* Decis. Support Syst. 54(2): 1104-1118 (2013)

- Tobias Mettler, Markus Eurich, Robert Winter: 
References


References

• Andreas Lanz, Barbara Weber, Manfred Reichert: 

• Clara Ayora, Victoria Torres, Barbara Weber, Manfred Reichert, Vicente Pelechano: 

• Barbara Weber, Manfred Reichert, Stefanie Rinderle-Ma: 

• Nick Russell, Wil M. P. van der Aalst, Arthur H. M. ter Hofstede: 

• Yair Wand, Ron Weber: 

• Stefan Zugal, Cornelia Haisiackl, Jakob Pinggera, Barbara Weber: 

Questions?