Evaluation of Design Science instantiation artifacts in Software engineering research*

Marko Mijač
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2, 42000 Varaždin, Croatia
{marko.mijac}@foi.hr

Abstract. *Design is a process of creating applicable solutions to a problem, and as such has long been accepted research paradigm in traditional engineering disciplines. More recently, it has been frequently used in the field of information systems and software engineering.*
One of the proposed approaches for conducting systematic and methodological design is Design Science (DS). It is essentially a pragmatic, problem-solving paradigm which results in development of construct, method, model or instantiation artifacts. However, in order to add science to Design Science, developed artifacts need to be properly evaluated.
In this paper we present guidelines for defining and performing evaluation of Design Science instantiation artifacts in software engineering research.

Keywords. Design Science, artifacts, evaluation, software engineering

1 Introduction

According to Merriam-Webster dictionary *design* indicates *planning and making something for a specific use or purpose*. As a process of creating applicable solutions to a problem, design has long been accepted research paradigm in traditional engineering disciplines. More recently, it has been frequently used in the field of information systems and software engineering.

One of the proposed approaches for conducting systematic and rigorous design is Design Science (DS). It is essentially a pragmatic, problem-solving paradigm which results in development of innovative artifacts, namely: constructs, methods, models and instantiations [1].

While each of these artifact types may appear as individual output of DS, proposed solution often consists of several artifacts being built upon one another. Instantiations are frequently at the top of such artifact stack, i.e. they are using domain constructs and implementing underlying models and methods. March et al. [1] describe instantiations as the realization of an artifact in its environment. In the context of software engineering research typical representatives of instantiations are implementations and prototypes of information systems, database systems, tools, components, services, libraries, frameworks, algorithms etc.

Apart from artifacts being innovative and relevant to a problem domain, in order to add *science* to Design Science developed artifacts need to be properly evaluated. Indeed, evaluation activities are present in each method, framework and guidelines for conducting design science research (DSR).

Due to difference in their purpose, form and characteristics, constructs, models, methods and instantiations as different artifact types

*This paper is published and available in Croatian language at: http://ceciis.foi.hr may require different approach to evaluation. Although our primary interest are instantiation artifacts, evaluation of instantiations and evaluation of constructs, models and methods embodied in instantiations is related in bidirectional manner. On one hand, evaluation of underlying constructs, models and methods certainly increases the overall quality of the resulting instantiation. On the other hand, as March et al. [1] report, by building instantiations we operationalize constructs, models and methods they contain, thus demonstrating their feasibility and effectiveness. Further, by evaluating instantiations we also provide confirmation for underlying artifacts.

In this paper we investigate existing methods, patterns, frameworks, and guidelines for performing evaluation in Design Science. We then proceed to extend the current state with our own guidelines for evaluation of Design Science instantiation artifacts in software engineering research.

The paper is structured as follows. Section 2 discusses DS evaluation in general and its position within existing DS research methods. In Section 3 the process of designing evaluation in DS research is discussed, with the emphasis on FEDS framework [2] and its potential extension with contributions from other papers. In section 4, we synthetize existing approaches and offer 7 high-level guidelines for designing and performing evaluation in design science. Finally, in section 5 we conclude the topic.

2 Evaluation in Design Science

2.1 Position of evaluation within Design Science process

Evaluation is the process of judging some thing's quality, importance, or value. Together with build activity, it constitutes the internal, build-evaluate design cycle, which, according to Hevner [3], is the heart of any design science research project.

A number of authors worked on formalizing the process of design science research. For example, in methodological framework proposed by Johannesson and Perjons [4], two out of five activities are dedicated to evaluation, namely *Demonstrate artefact* and *Evaluate artefact*. Demonstration can here be considered as weak form of evaluation, and it shows that artifact is feasible and that it works. Evaluation activity on the other hand aims to examine how well the artifact works. Similar proposal comes from Peffers et al [5]. In their design science process model two steps are specified as *Demonstration* and *Evaluation*. Vaishnavi et al. [6] propose a general methodology of DSR with one of the phases being *evaluation*. Wieringa [7] puts the design science into a perspective of engineering cycle, and proposes validation and evaluation activities. He describes validation as a means to predict how artifact will interact with its context, prior to building artifact. On the other hand, evaluation investigates how implemented performs in real-world context. Evaluation is represented also as a guideline in well-known design science guidelines from Hevner et al. [8]. Offermann et al. [9] offer formalization of detailed DSR process, with one of the three phases being evaluation phase. Sein et al. [10] placed evaluation in second stage (Building, Intervention, and Evaluation) of their Action Design Research method.

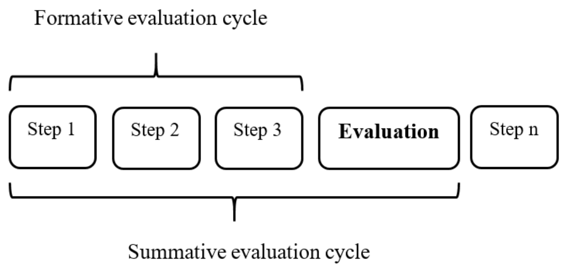
2.2 Evaluation cycles

As can be seen, evaluation is an inherent part of every formal process of design science research. Most approaches depict evaluation as clearly separated phase or step which is performed after artifact is designed and built. However, design science process is not necessarily performed as a waterfall model, but can contain iterations and cycles. For example, the outputs from evaluation activity can result in going back to previous phases by uncovering the flaws in artifact's design and build, altering understanding of the initial problem, or simply yielding ideas for new and improved design. Although their framework looks sequential, Johannesson and Perjons [4] support this iterative style by stating that design science is always carried out in iterative way, moving back and forth between all activities. Offermann et al. [9] also emphasize that depending on results of evaluation phase, one can iterate back to previous phases.

Sein et al. [10] go further in their Action Design Research method, and claim that evaluation activity is inherently interwoven with building the artifact and intervening in organization, and that they should be carried out concurrently. When discussing the purpose of evaluation Venable et al. [11] distinguish formative and summative evaluation. Formative evaluations focus on providing feedback and measuring improvement as development progresses, while summative evaluation supports forming opinion about artifact and comparing artifact after development is completed.

In general, regardless of the chosen design science process we can identify two evaluation cycles, namely formative and summative cycle. In formative evaluation cycle evaluation is carried out continuously and in parallel with designing and building artifact. It aims to provide feedback as early as possible in order to incrementally improve and refine artifact. In this cycle, possibly large number of iterations with implicit and explicit micro evaluations take place. Summative evaluation cycle, on the other hand, assumes that artifact has been built and explicit formal evaluation of artifact as a final result of design science research can start. In this evaluation step the artifact could also be marked as unsatisfactory, and it could be required to go back to previous steps and to improve the artifact. However, the number of iterations in summative evaluation cycle is usually much smaller. Here, it is important to note, that evaluation will seldom conclude that the evaluated artifact is perfect and that no improvements are possible. Therefore, researcher should keep in mind the goals and the limitations of the research project, and estimate when iterations and improvements should stop, or at least be deferred to future research.

Figure 1 Evaluation cycles in Design Science research process



2.3 Instantiations

Gregor and Jones [12] describe instantiations as material artifacts which have physical existence in the real world, and are fundamentally different from constructs, models and methods, described as abstract artifacts. March et al. [1] indicate that instantiation is the realization of an artefact in its environment. Similarly, Johannesson and Perjons [4] describe instantiation as a working system that can be used in practice.

Instantiations can also be characterized in terms of the difference between *product artifacts* and *process artifacts* [11]. While process artifacts represent methods and procedures which guide people in accomplishing some task, product artifacts represent tools, diagrams, software etc., which people use to accomplish some task. Evidently instantiation artifacts in software engineering will in most cases appear as product artifacts.

Another view on instantiation artifacts in software engineering is from the perspective of *technical artifacts* and *socio-technical artifacts* [11]. In that sense, most instantiations in software engineering appear in the form of socio-technical artifacts, meaning they are technical systems but are required to interact with humans to be useful (e.g. information and ERP systems, games, CASE tools, etc). On the other hand, instantiations can also appear as purely or predominantly technical artifacts, which means they require no or minimum of interaction with humans (e.g. software components embedded into larger, possibly socio-technical artifact).

From the perspective of evaluation in design science, instantiations are particularly important. For example, March et al. [1] claim that instantiations operationalize embedded constructs, models, and methods, thereby demonstrating their feasibility and effectiveness. Similarly, according to Hevner et al. [8] instantiations show that constructs, models, or methods can be implemented into a working system. They demonstrate feasibility, enabling concrete assessment of an artifact's suitability to its intended purpose. Gregor and Jones [12] conclude that while conceptual work on design has proved to be influential in computing, the credibility of work is likely to be enhanced by providing instantiation as a working example.

3 Design of evaluation

While all design science research processes address evaluation and offer general hints and tips on how to conduct it, they lack the exact and detailed evaluation procedure. Johannesson and Perjons [4] indicate that the very use of scientific research methods in performing evaluation is the key to differentiate design science from routine design. Venable et al. [2] add that, if design science research is to deserve its label as "science", the evaluation should be relevant, rigorous, and scientific.

In order to plan, design and perform such rigorous evaluation activities, appropriate procedures, frameworks, guidelines are needed. Depending on the characteristics of the design science research project and the concrete artifact, they should aid us in deciding when, what, why and how to evaluate.

3.1 Existing approaches

Recently, a number of authors addressed the problem of designing evaluation within DSR. Pries-Heje [13] propose a strategic framework for DSR evaluation, which can be used both to aid in selecting appropriate evaluation strategy for novel research, as well as to classify evaluation strategies in already published research. The framework is based on two dimensions: *ex-ante vs ex-post* evaluations, and *naturalistic vs artificial* evaluation. Cleven et al. [14] present a morphological field with 12 variables and their respective values, which can be used to decide design alternatives for evaluation strategy. Venable et al. [2] developed a Framework for Evaluation in Design Science research (FEDS), which specifies four-step procedure for designing evaluation strategy. Sonnenberg and Brocke [15] present general design science research evaluation pattern which prescribes four evaluation activities to be carried-out through entire DSR process.

3.2 Chosen approach

In order to discuss specifics of designing evaluation, we will rely on FEDS framework proposed by Venable et al. [2]. The framework specifies four-step procedure for designing evaluation: (1) *explicate the goals of evaluation*, (2) *choose the evaluation strategy*, (3) *determine the properties to evaluate*, (4) *design the individual evaluation episode*. However, since FEDS framework does not consider evaluation criteria systematically, nor does it relate them to evaluation methods, we will complement steps (3) and (4) with findings from other relevant research.

3.2.1 Explicate the goals of evaluation

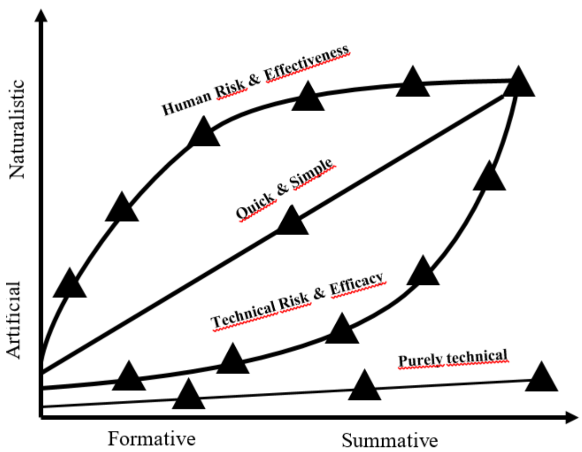
Venable et al. [2] name four competing goals which we must consider when designing evaluation: (1) rigour, (2) uncertainty and risk reduction, (3) ethics and (4) efficiency. *Rigour* is considered here in terms of: efficacy (establishing that improvements are really caused by the artifact) and effectiveness (establishing that artifact works in real situations). Note however that formative evaluation cycle is more appropriate for evaluating efficacy, and summative evaluation cycle for evaluating effectiveness. *Uncertainty and risk reduction* considers effort to eliminate or reduce social and technical risks as early as possible. Formative

evaluation by definition is particularly important for this goal.
Ethical issues should be attended especially when evaluating artifacts which introduce the safety, health and privacy risks.
Finaly, the evaluation should be *efficient* in terms of being feasible within the limited research resources (time, money, people…).

3.2.2 Choose the evaluation strategy

In order to characterize and position different evaluation strategies, Venable et al. [2] propose two dimensional space, with dimensions being (1) functional purpose and (2) paradigm of the evaluation. The functional purpose dimension addresses the question: why to evaluate; and positions evaluation according to aforementioned formative-summative evaluation continuum. Paradigm of the evaluation addresses the question: how to evaluate; and forms artificial vs naturalistic evaluation continuum. As the name itself implies, artificial evaluation is carried out in artificial (e.g. laboratory or simulator) environment, while naturalistic evaluation is carried in artifact's real or as real as possible environment and under realistic conditions.

Figure 2 Two-dimensional space for evaluation strategies [2]



In this two-dimensional space, evaluation strategy is represented as a trajectory formed by connecting individual evaluation episodes. The evaluation episodes were previously positioned according to two dimensions. Venable et al. [2] propose four archetypes of evaluation strategy, namely: Human risk & Effectiveness, Quick & Simple, Technical risk & Efficacy, and Purely technical. With those archetypes a simple heuristics is provided to help researchers pick the most appropriate archetype for their research. Authors, however, emphasize that each design science research is specific, and encourage researchers to adapt proposed archetypes if necessary, or even to propose new evaluation strategies.

3.2.3 Determine the properties to evaluate

In order to determine what exact properties of instantiations to evaluate, we need to consider a number of criteria, including the general goals of evaluation, chosen strategy, characteristics of artifact we are evaluating, artifact's purpose. According to Venable et al. [2] the final selection of properties is necessarily unique to the artifact.
Different authors proposed different properties/criteria to evaluate instantiations. March et al. [1], for example, consider *efficiency*, *effectiveness*, and the artifact's *impact on the environment and users*. Hevner et al. [8] in their evaluation guideline state that artifact must demonstrate *utility*, *quality*, and *efficacy*.
While no commonly accepted list of evaluation properties exist, Prat et al. [16] analyzed design science literature and reported the list of evaluation properties with their occurrence frequency. This list provides a good start point for choosing evaluation properties which are appropriate for particular artifact. However, one should also keep in mind the frequency of properties in literature, because higher frequency may indicate already established best practice and possibly better acceptance from the reviewers.
From the original list of properties [16] we excluded *construct deficit*, because it obviously refers to evaluation of construct artifacts. Other than that, we argue that the properties in **Table 1** can be applied when evaluating instantiation artifacts.

Table 1 Occurrence frequency (column f) of evaluation properties in DS research (adapted from Prat et al. [16])

Criteria	Description	f (%)
Efficacy	The degree to which the artifact achieves its goal considered narrowly, without addressing situational concerns.	80%
Usefulness	The degree to which the artifact positively impacts the task performance of individuals.	35%
Technical feasibility	Evaluates, from a technical point of view, the ease with which a proposed artifact will be built and operated.	32%
Accuracy	The degree of agreement between outputs of the artifact and the expected outputs.	28%
Performance	The degree to which the artifact accomplishes its functions within given constraints of time or space.	23%
Effectiveness	The degree to which the artifact achieves its goal in a real situation.	18%
Ease of use	The degree to which the use of the artifact by individuals is free of effort.	10%
Robustness	The ability of the artifact to handle invalid inputs or stressful environmental conditions.	10%
Scalability	The ability of the artifact to either handle growing amounts of work in a graceful manner, or to be readily enlarged.	10%
Operational feasibility	Evaluates the degree to which management, employees, and other stakeholders, will support the proposed artifact, operate it, and integrate it into their daily practice.	10%
Utility	Utility measures the value of achieving the artifact's goal, i.e. the difference between the worth of achieving this goal and the price paid for achieving it.	7%
Validity	Validity means that the artifact works correctly, i.e. correctly achieves its goal.	6%
Completeness	The degree to which the activity of the artifact contains all necessary elements and relationships between elements.	3%
Adaptability	The ease with which the artifact can work in contexts other than those for which it was specifically designed. Synonym: flexibility	2%
Reliability	The ability of the artifact to function correctly in a given environment during a specified period of time.	2%
Learning capability	The ability of the artifact to learn from experience.	2%
Simplicity	The degree to which the structure of the artifact contains the minimal number of elements and relationships between elements.	1%
Economic feasibility	Evaluates whether the benefits of a proposed artifact would outweigh the costs of building and operating the artifact.	1%
Generality	Refers to the scope of the artifact's goal. The broader the goal scope, the more general the artifact.	1%

3.2.4 Design individual evaluation episodes

Evaluation episode Ep can be specified as a concrete evaluation within evaluation strategy, characterized by 4 dimensions: evaluation purpose (Pu), evaluation paradigm (Pa), evaluation method (M) and one or more evaluation properties (Pr):

$\boldsymbol{Ep = \{Pu, Pa, M, Pr(p1, p2, \dots)\}}$, where
$Pu = (formative \mid summative)$,
$Pa = (artificial \mid naturalistic)$,
$Pr = (efficacy \mid usefulness \mid \dots)$,
$M = (experimentation \mid case\ study \mid \dots)$,

While we already discussed evaluation purpose, paradigm and properties, the question of potential evaluation methods remains. **Table 2** shows some of the most frequently mentioned evaluation methods and patterns in design science literature.

Some authors also performed literature review of DS research papers in order to found out what evaluation methods are really used by researchers. For example, Peffers et al. [17] reviewed 148 DS research articles and reported technical experiment, subject-based experiment, prototyping, and demonstration through illustrative scenarios to be dominant evaluation methods for instantiation artifacts. Prat et al. [16] developed a taxonomy of evaluation methods by examining 121 DS research papers. They identified demonstration (on illustrative or real examples), simulation and benchmarking, case study, and controlled experiment, as most represented evaluation techniques for instantiation artifacts.

Table 2 Evaluation methods and patterns in DS

Evaluation method/pattern	Mentioned in
Experimentation	[18][2][8][17][6][9][7][16][14]
Case study	[18][2][8][17][9][7][16][14]
Simulation	[18][2][8][6][9][7][16]
Informed argument	[18][2][8][17][6][15]
Demonstration / Scenarios	[18][8][17][6][15][16]
Field study	[18][2][8][14]
Mathematical proofs	[18][2][6][14]
Survey	[18][2][7][14]
Action research	[2][17][9][14]
Expert evaluation	[18][17][9]
Benchmarking	[18][6][16]
Static/Dynamic analysis	[18][17][16]
Prototyping	[17][15][14]
Testing	[8][7]
Metrics	[2][16]

Prat et al. [16] also identified six commonly used compositional styles for evaluation of instantiation artifacts, namely: *(1) demonstration, (2) simulation and metric-based benchmarking, (3) practice-based evaluation of effectiveness, (4) simulation and metric-based absolute evaluation, (5) practice-based evaluation of usefulness or ease of use, and (6) laboratory, student-based evaluation of usefulness.* When building evaluation strategy, these compositional styles can be used as already established evaluation episodes.

4 Guidelines on designing and performing evaluation of instantiations

Evaluation has been acknowledged as one of the key activities in design science research. This can be seen from research papers dealing with design science theory, as well as from research papers conducting DSR. In order to conduct design science evaluation in a systematic and rigorous way we present several guidelines synthetized from design science literature.

Guideline 1 – *Use established frameworks for design science research*

Ignoring the very importance of choosing relevant research problem, the first step a researcher doing design science research can do in terms of evaluation is to choose appropriate design science method/process. While choosing and following a good method/process is not necessarily a guarantee of producing good artifact, it definitely increases a chance of it happening. In addition, every design science research method incorporates evaluation step and positions evaluation with regard to other design science activities. Examples of formalized methods for conducting DSR can be found in [4], [5], [6], [7], [8], [9], [10].

Guideline 2 – *Use existing frameworks for design of evaluation*

After general design science method is chosen and positioned, the next evaluation-related activity is to design evaluation. Designing evaluation is complex task and, same as the design science process itself, it needs to be conducted systematically. In order to do that, researcher can follow one or combination of existing approaches reported in section 3.1., namely: [2], [13], [14], [15]. However, in our opinion FEDS framework [2], with its four-step procedure, currently offers the most comprehensive guidance.

Guideline 3 – *Consider evaluating commonly evaluated artifact properties when designing evaluation*

When determining what artifact properties to evaluate one should consult papers from guideline 2. For example, FEDS framework [2] offers heuristics for this step. However, researcher should consider consulting **Table 1** which provides a source of frequently evaluated properties in design science research. Frequently evaluated properties may indicate best practice and possibly better acceptance from reviewers.

Guideline 4 – *Consider commonly used evaluation methods when designing evaluation*

While nothing prevents a researcher to choose whatever method he or she finds suitable for evaluating particular artifact property, it is useful to consider those which are commonly used. **Table 2** contains evaluation methods and patterns which are frequently mentioned as potential evaluation methods throughout design science literature. Also, papers [16] and [17] report evaluation methods most frequently applied in design science research articles.

Guideline 5 – *Consider commonly used evaluation compositional styles when designing evaluation*

Designing concrete evaluation episodes within overall evaluation strategy includes determining which particular method will be used to evaluate particular artifact properties. While large number of property-method combinations can be formed by pairing each and every evaluation property and method, some of these combinations are more common than others. Common evaluation compositional styles reported in [16] can be consulted when deciding on evaluation strategy and individual evaluation episodes.

Guideline 6 – *Use appropriate frameworks for performing particular evaluation methods*

Using particular evaluation method in evaluation episodes is often research within research. Research methods used as evaluation methods in design science have precisely defined steps on how to conduct them, e.g. planning, collecting data, analyzing data and reporting. One should consider finding and using frameworks or methods on performing particular evaluation method within design science, if such exist. For example, following methods are discussed in the context of design science evaluation: Focus groups [19], Software embedded evaluation [20], Technical action research [21], and Experimentation [22][23]. Alternatively, frameworks and methods discussed in the context of e.g. software engineering or other fields may be perfectly suitable as well: Case study [24], Experimentation [25], Action research [26] etc.

Guideline 7 – *Consider using established software quality models and metrics to evaluate instantiations*

Various quality models have been proposed in order to assess the quality of software products, one of them being ISO/IEC 25010:2011 standard [27]. This quality model, for example, prescribes eight quality characteristics (subdivided into sub characteristics) together with corresponding quality measures and functions used for quantifying those characteristics. According to Pries-Heje et al. [13] when evaluated artifact is a product, we can use established software quality models in terms of evaluation.

5 Conclusion

In this paper we discussed evaluation of instantiation artefacts in DS research. This poses a significant undertaking and often entire new research within the DS research. In order to aid in designing and performing systematic and rigorous evaluation, we offered 7 guidelines. The guidelines are high-level in terms that they do not deal with performing specific evaluation methods or criteria. Rather, they guide researcher towards existing frameworks and methods for positioning evaluation in their DS research, designing evaluation, choosing established evaluation properties and methods. Although the paper is focused on evaluation of instantiation artifacts, the guidelines in greater part are applicable also to other artifact types.

References

[1] S. T. March and G. F. Smith, “Design and natural science research on information technology,” *Decis. Support Syst.*, vol. 15, no. 4, pp. 251–266, Dec. 1995.

[2] J. Venable, J. Pries-Heje, and R. Baskerville, “FEDS: a Framework for Evaluation in Design Science Research,” *Eur. J. Inf. Syst.*, Studeni 2014.

[3] A. Hevner, “A Three Cycle View of Design Science Research,” *Scand. J. Inf. Syst.*, vol. 19, no. 2, Jan. 2007.

[4] P. Johannesson and E. Perjons, *An introduction to design science*. 2014.

[5] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, “A Design Science Research Methodology for Information Systems Research,” *J. Manag. Inf. Syst.*, vol. 24, no. 3, pp. 45–77, Dec. 2007.

[6] V. Vaishnavi, *Design science research methods and patterns: innovating information and communication technology*. Boca Raton: Auerbach Publications, 2008.

[7] R. J. Wieringa, *Design Science Methodology for Information Systems and Software Engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.

[8] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design science in information systems research,” *MIS Q.*, vol. 28, no. 1, pp. 75–105, 2004.

[9] P. Offermann, O. Levina, M. Schönherr, and U. Bub, “Outline of a design science research process,” 2009, p. 1.

[10] M. K. Sein, O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren, “Action Design Research,” *MIS Q*, vol. 35, no. 1, pp. 37–56, Ožujak 2011.

[11] J. Venable, J. Pries-Heje, and R. Baskerville, “A Comprehensive Framework for Evaluation in Design Science Research,” in *Design Science Research in Information Systems. Advances in Theory and Practice*, K. Peffers, M. Rothenberger, and B. Kuechler, Eds. Springer Berlin Heidelberg, 2012, pp. 423–438.

[12] S. Gregor and D. Jones, “The Anatomy of a Design Theory,” *J. Assoc. Inf. Syst. Atlanta*, vol. 8, no. 5, pp. 312-323,325-335, May 2007.

[13] J. Pries-Heje, R. Baskerville, and J. Venable, “Strategies for Design Science Research Evaluation,” *ECIS 2008 Proc.*, Jan. 2008.

[14] A. Cleven, P. Gubler, and K. M. Hüner, “Design Alternatives for the Evaluation of Design Science Research Artifacts,” in *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*, New York, NY, USA, 2009, pp. 19:1–19:8.

[15] C. Sonnenberg and J. vom Brocke, “Evaluation patterns for design science research artefacts,” in *Practical Aspects of Design Science*, Springer, 2011, pp. 71–83.

[16] N. Prat, I. Comyn-Wattiau, and J. Akoka, “A Taxonomy of Evaluation Methods for Information Systems Artifacts,” *J. Manag. Inf. Syst.*, vol. 32, no. 3, pp. 229–267, Jul. 2015.

[17] K. Peffers, M. Rothenberger, T. Tuunanen, and R. Vaezi, “Design Science Research Evaluation,” in *Design Science Research in Information Systems. Advances in Theory and Practice*, K. Peffers, M. Rothenberger, and B. Kuechler, Eds. Springer Berlin Heidelberg, 2012, pp. 398–410.

[18] C. Sonnenberg and J. vom Brocke, “Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research,” in *Design Science Research in Information Systems. Advances in Theory and Practice*, vol. 7286, K. Peffers, M. Rothenberger, and B. Kuechler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 381–397.

[19] M. Tremblay, A. Hevner, and D. Berndt, “Focus Groups for Artifact Refinement and Evaluation in Design Research,” *Commun. Assoc. Inf. Syst.*, vol. 26, no. 1, Jun. 2010.

[20] L. Chandra Kruse and et al., “Software Embedded Evaluation Support in Design Science Research,” presented at the Pre-ICIS workshop on Practice-based Design and Innovation of Digital Artifacts, 2016.

[21] R. Wieringa and A. Morali, “Technical Action Research as a Validation Method in Information Systems Design Science,” in *Design Science Research in Information Systems. Advances in Theory and Practice*, K. Peffers, M. Rothenberger, and B. Kuechler, Eds. Springer Berlin Heidelberg, 2012, pp. 220–238.

[22] L. Ostrowski and M. Helfert, *Design Science Evaluation – Example of Experimental Design*. .

[23] T. Mettler, M. Eurich, and R. Winter, “On the Use of Experiments in Design Science Research: A Proposition of an Evaluation Framework,” *Commun. Assoc. Inf. Syst.*, vol. 34, no. 1, Jan. 2014.

[24] B. Kitchenham, L. Pickard, and S. L. Pfleeger, “Case studies for method and tool evaluation,” *IEEE Softw.*, vol. 12, no. 4, pp. 52–62, Jul. 1995.

[25] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[26] D. E. Avison, F. Lau, M. D. Myers, and P. A. Nielsen, “Action research,” *Commun. ACM*, vol. 42, no. 1, pp. 94–97, Jan. 1999.

[27] ISO, “ISO/IEC 25010:2011 - Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models.” 2011.