AIS SIGPRAG

Pre-ICIS Workshop 2016
"Practice-based Design
and Innovation of
Digital Artifacts"

Evaluation of Design Artifacts - more than prototypes and case studies

Patrick Brandtner [(1)], Markus Helfert [(2)], Andreas Auinger [(1)], Kurt Gaubinger [(1)]

[(1)] University of Applied Sciences Upper Austria

[(2)] Dublin City University

Abstract

Evaluation of design artefacts is of crucial importance in design science research (DSR). A plethora of evaluation approaches and methods can be found in literature; nevertheless, little work has been done so far to investigate the relation between the evaluation strategies, methods and techniques in DSR evaluations. Prototype implementations, together with case studies seem to be dominant and the technique of choice to evaluate, often complex artefacts. This paper goes beyond the common approach in DSR, and presents a multi-media and web-based DSR evaluation approach. We present the definition of evaluation criteria, the selection of evaluation methods and the findings and experiences gained. The results of this paper can support other design science research approaches concerned with the evaluation of concepts or process models.

Keywords: Design science research, artefact evaluation, conceptual modelling, process model

1 Introduction and Background

In design science research (DSR) artefact evaluation within a specific environment is of crucial importance (Goldkuhl 2013; Peffers et al. 2012). Artefacts should be evaluated based on the requirements of the context of their respective application and implementation environment (Peffers et al. 2012). According to Niederman et al. (2012) the initial evaluation of a novel artefact may simply be to show that it works and produces adequate solutions. The challenge is to define "adequacy". However, as evaluation criteria are socially constructed, what one researches may consider adequate, may be considered inadequate by another (Niederman, March 2012). Same applies to the selection of a suitable evaluation method.

A plethora of evaluation approaches and methods in DSR can be found in literature, e.g. in (Peffers et al. 2012; March, Smith 1995; Helfert et al. 2012). Nevertheless, little work has been done so far specifically addressing the choice and combination of evaluation strategies, methods and tools in DSR evaluations (Prat et al. 2014; Sonnenberg, Brocke 2012). Numerous design science evaluation approaches centre on prototype implementation and build the analysis around usability. They often apply the well-known method of case study research (Peffers et al. 2007; Markus et al.

2002). However, DSR evaluation can go far beyond prototypes, case studies and interviews.

This paper aims at presenting a multi-media, web-bassed DSR evaluation approach in the form of a survey, enriched with multi-media content. The evaluation approach described in this paper is part of a longer term project that includes the development of a process model for innovation management. More specifically, the aim of this project is to develop a process model specifically addressing the Front End of Innovation, which represent the earliest stages of innovation and includes the activities that come before the formal and well-structured new product development (NPD) process. The process model is described using Event-driven-process-chains (EPC) and has been evaluated from an ex-ante and an ex-post perspective. The study context and the ex-ante aspects of the evalation using focus group research have been document in Brandtner et al. 2015. Expanding our earlier work, this paper focuses on the ex-post evaluation and how a multi-media, web-based appraoch was employed.

The remainder of this paper is structured as follows: In the subsequent sections, a short overview of literature on DSR evaluation (section 2), the ex-post evaluation approach of the research project mentioned above (section 3) and the findings and learnings of the application and conduction of this evluation approach are discussed and explained (section 4).

2 Evaluation in DSR

Even though the importance of artefact evaluation is acknowledged in many design science contributions, many researchers focus often on a fragmented or incomplete lists of evaluation criteria, with often emphasising utility and usability. However, the application of appropriate evaluation criteria is essential in scientific research in general and in particular in design science projects where artefacts have to be assessed against criteria of value or utility (March, Smith 1995). In addition, often evaluation methods are described without guidance on how to apply which methods to which criteria (Prat et al. 2014; Ostrowski, Helfert 2012).

One of the few papers addressing the definition of DSR evaluation strategies was contributed by Pries-Heje et al., who proposed a framework supporting researchers in building such strategies. They distinguish between three core dimensions of an evaluation strategy: when to evaluate, what to evaluate and how to evaluate (Pries-Heje et al. 2008). “When to evaluate” aims at defining if an ex-ante or ex-post evaluation is needed. Regarding the “what to evaluate” perspective, the objective is to define whether to evaluate artefact design process or the artefact design product (Sonnenberg, Brocke 2012). Finally, “how to evaluate” relates to the form of evaluation and may be naturalistic or artificial. Naturalistic evaluation focusses on exploring respectively evaluating the artefact in its real environment, in our instance in the organisations of survey participants (Venable et al. 2012). Artificial evaluation on the other hand includes laboratory settings, field experiments, mathematical proof or simulations.

3 Developing a multi-media DSR Evaluation approach

The following sections describes the development of a multi-media DSR evaluation approach together with a discussion on evaluation criteria, the selection of appropriate evaluation methods and the combination of these into a coherent evaluation approach.

3.1 Selection and measurement of evaluation criteria

Many researchers emphasis the importance of utility, however utility of artefacts is complex and may depend on various attributes, its use or the artefact itself (Ostrowski, Helfert 2012). In addition, the term utility is often used synonymously to the term usefulness in literature (Prat et al. 2014) and utility has often been assessed through e.g. perceived usefulness (Adipat et al. 2011, Reeder et al. 2011). Therefore, we prefer to use the term usefulness or perceived usefulness rather than the term utility.

Content measures for artefact evaluation are often closely linked with quality criteria, as quality can be described in terms of more or less measurable sets of criteria. Differences in quality measurement results reflect differences in the state or quantity of specific artefact attributes (Pries-Heje et al. 2008). Many different definitions of quality can be found in literature. The underlying assumption of process-based quality is that a good process will lead to a good process outcome respectively result or product (Pries-Heje et al. 2008). In terms of process model quality respectively of quality in conceptual modelling, many approaches (e.g., Rittgen 2010, Venkatesh, Davis 2000, Helfert et al. 2012) focus on three core levels of quality, which will be used as the three main evaluation dimensions of our ex-post process model evaluation approach: Syntactic Quality (SNQ), Semantic Quality (SMQ) or Perceived Semantic Quality (PSQ), and Pragmatic Quality (PMQ).

Syntactic process model quality (SNQ)

The syntactic quality of a model refers to the extent to which it observes the rule of its underlying modelling language (Rittgen 2010). A model is correct from a syntactical point of view if all statements of the model are according to the syntax and vocabulary of the modelling language respectively the underlying notation. An EPC-process model has to fulfil certain criteria to be sound. A number of approaches that used modelling conventions as a metric for syntactic quality can be found in literature (Rittgen 2010).

Perceived semantic process model quality (PSQ)

Semantic quality measures model quality in terms of what the model includes that is not present in the domain and of what the model does not include that is present in its domain (Liu et al. 2012). Maes and Poels proposed and validated a four-indicator measurement system including correctness, completeness, authenticity (realistic) and relevance (Rittgen 2010; Maes, Poels 2007). For evaluating the Perceived Semantic Quality (PSQ) of our process model, we follow this four-indicator PSQ-system. The concrete statements for measuring PSQ were taken from the validated PSQ-measurement system of Rittgen (Rittgen 2010). All items are measured on a 7-point Likert scale, where 1 - strongly disagree, 2 - moderately disagree, 3 - somewhat disagree, 4 - neutral (neither disagree nor agree), 5 - somewhat agree, 6 - moderately agree, and 7 - strongly agree.

Pragmatic process model quality / Perceived process model usefulness (PU)

Pragmatic process model quality describes a process model's ability or usefulness in order to facilitate learning and action in an organisational context (Maes, Poels 2007; Krogstie et al. 2006). Applied to the current research project, pragmatic process model quality hence describes the usefulness of the model in real organisational Front End of Innovation processes. In the context of our evaluation approach, pragmatic process model quality will be measured based on the perceived usefulness of our process model as rated by real users. The concrete statements for measuring PU were adapted to the current research background based on the original statements of Venkatesh and Davis (Venkatesh, Davis 2000) and were – like for PSQ - measured on a 7-point Likert scale.

The process model quality dimensions of our evaluation approach range from syntactic model quality (modelling notation,), to semantic quality (domain knowledge) and to pragmatic model quality (perceived usefulness of the model in its application domain). Syntactical issues are well controlled and can be measured objectively. The main evaluation effort would therefore be directed towards semantic and pragmatic model quality, which are potentially harder to measure and evaluate (Krogstie et al. 2006). Figure 1 provides an overview of the quality dimensions and their respective items as discussed above.

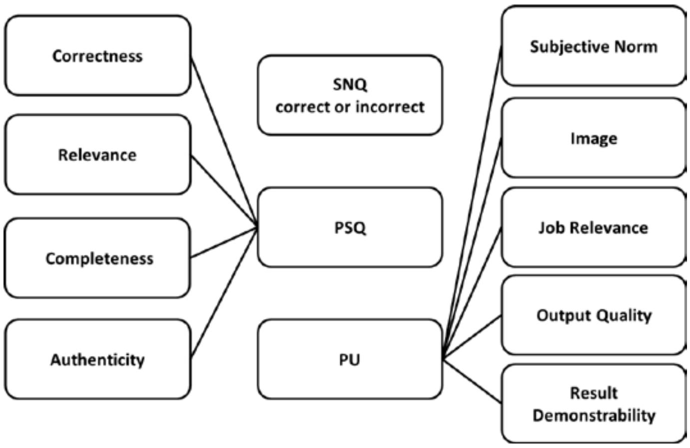

Figure 1: Summary of quality dimensions of our process model and their respective items

3.2 Design of multi-media evaluation tool

For the evaluation, we selected a semi-quantitative, questionnaire based survey with qualitative comment fields for data gathering and collection (e.g. Venable et al. 2012). In contrast of a typical case study or prototype evaluation, we argue the following study design has some advantages to evaluate process models:

The actual questionnaire is preceded by an introduction to the process model. Therefore, we created one introduction video presentation for all participants presented via a web-based survey tool. This allowed us to reduce bias caused by different and varying forms of process model presentation and different accompanying explanations to its modules and activities. We developed the introduction process model presentation using MS PowerPoint. Based on the single slides of this presentation, we elaborated a comment guideline for each part. The comments where then recorded and the audio files generated this way where matched to the respective slides and where included into the final PowerPoint file. Additionally, we used the pointer-feature of PowerPoint to highlight the relevant parts of the single slides according to the timing and content of the audio comment files. The presentation was then saved as a video file, which was uploaded to YouTube (http://www.youtube.com) in order to be integrated into our SoSci Survey online questionnaire. Figure 2 illustrates the survey layout and provides an overview of the introductory page of the survey including the video presentation.

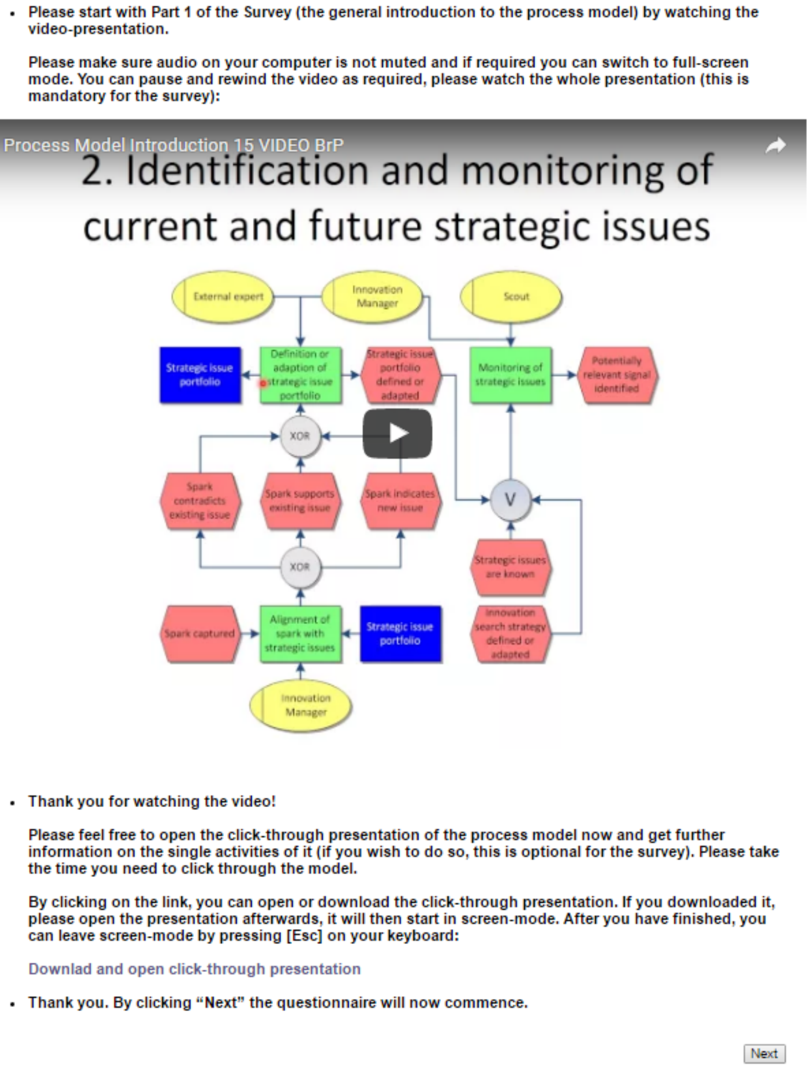

Figure 2: Layout of the questionnaire and integration of the video presentation

Following the introduction video presentation, the process model is explained in the form of a click-through and wiki-like presentation, allowing survey participants to get more detailed information about the process model.

The survey then consists of 9 items and 22 statements and additional textual statements. In order to gain additional feedback and qualitative input for process model discussion, textual comments are collected for the selected items of PSQ relevance, PSQ completeness, PU job relevance and PU result demonstrability. Comment fields are shown based on triggering answer options of participants, providing us with the possibility of considering qualitative aspects as well. Furthermore, the collection of textual feedback allows for further interpretations of survey results and ultimately provides us with the possibility to gain further learnings and insights.

The statements investigating the defined items were presented pairwise, the answer options are illustrated in figure 3:

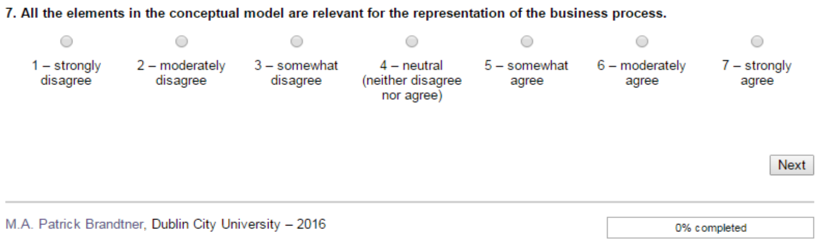

Figure 3: Layout of the questionnaire regarding the statements of PSQ and PU

As a survey tool for data collection, we selected SoSci Survey (https://www.soscisurvey.de), which allows for creating online questionnaires and for integration of additional media files. Microsoft Excel was used for data analysis. Web-based surveys allow for a quick, simple, cheap and effective way to reach a large population of potential participants and have been applied in a variety of settings and with many different populations (Brown et al. 2016). Surveys in general, and web-based surveys in particular represent a good evaluation technique for design methods and conceptual models, especially if the objective is to gather perception information from practitioners (Siau, Rossi 2011). Furthermore, survey and questionnaire design, dissemination and data storage and analysis are efficient and well supported by different survey tools (Greenlaw, Brown-Welty 2009).

In summary, the survey consisted of the following elements: Part 1: Introductionary presentation of the process model (10-minute video, mandatory); Part 2: Additional, information about the process model (optional click-through presentation, duration as required); Part 3: Completion of PSQ and PU questionnaire (10 minutes, mandatory).

4 Experiences of multi-media evaluation

4.1 Test Setting

As target groups for the survey, middle and executive management-level domain experts respectively actual users and beneficiaries of the process models were approached. Participants were invited by e-mail, the selection of potential respondents - complying with the requirements stated above – was distributed via two innovation management related organisations – namely the Platform of Innovation Management (PFI) (http://www.pfi.or.at/) and the Product Development and Management Association (PDMA) (http://www.inknowaction.com/pdma).

A pre-test survey with participants form academia was carried out in order to validate the general structure of the questionnaire, the performance and suitability of the survey tool, the measurement scale proposed and the textual comment functionality of the survey. The collected pre-test comments confirmed the design and structure of the questionnaire and the survey tool.

4.2 Findings and Discussion

The survey was open for participation from August to October 2016. In total, 52 participants from different industries (ranging from manufacturing, automotive, telecommunication and energy to IT services, construction, software, biotechnology etc.) and 5 five different countries completed our survey. The results of the survey indicated the quality of our process model and are summarised in figure 4:

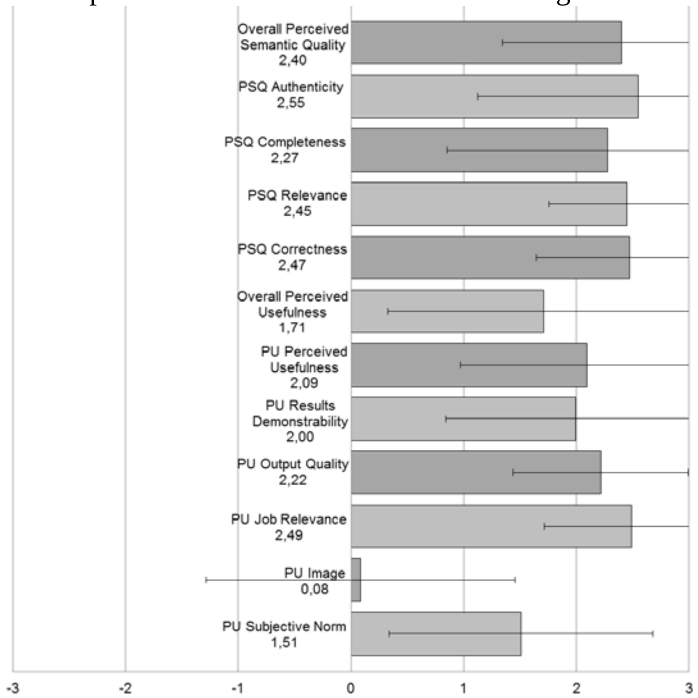

Figure 4: Average and standard deviation of PSQ, PU and their items (normalised, ranging from strongly disagree (-3) to strongly agree (3)

The conduction of the ex-post evaluation of our process model led to some important findings and key learnings:

(1) An important finding is the benefit of including and planning participant recruitment at an early stage of evaluation approach planning. In the context of this study, we aimed to address experts from the field of innovation and product management, business development or strategic planning. If experts are to be reached by the survey or are to be included in the evaluation approach, we recommend contacting and collaborating with respective groups, organisations or associations in order to address a large population of potential participants. In our context, participants were approached via the Platform of Innovation Management (PFI) and the Product Development and Management Association (PDMA). Additionally, we directly contacted past and current research partners of our university. In this context, we also recommend to contact potential participants (if contact details are available) once at the start of the survey and a second time one or two weeks before the end of it. Looking at the response rates, we could clearly identify the timing and the effect of the second notification respectively reminder mail which we sent out on September 19th:

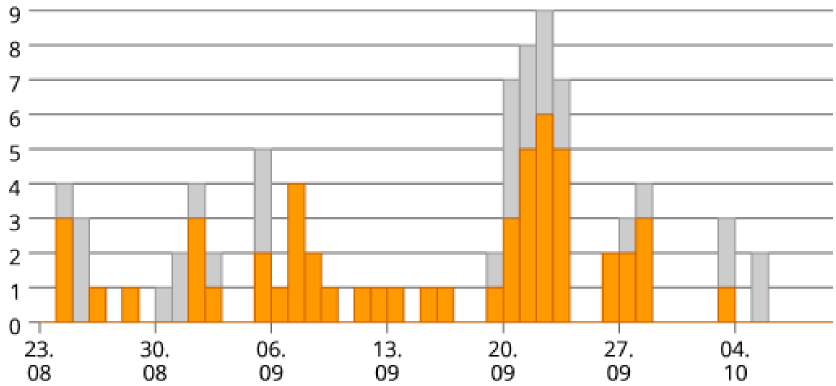

Figure 5: Responses over time

(2) Another key learning is to keep in mind the total duration of the evaluation and the way of how to communicate duration and timing related aspects to participants. In our case, the web-based survey took participants between 20 to 30 minutes to complete the mandatory parts of the survey, including 10 minutes of watching the video presentation and between 10-15 minutes of filling out the actual questionnaire. Additional time was necessary and further increased the duration of the survey if participants chose to click through the screen-mode wiki presentation after the video. The participant invitations sent out in the course of participant recruitment also included - besides some general survey and contact information - an overview of the single mandatory and optional parts of the survey and the time estimated to complete these parts. Although we tried to keep the introductory video as short as possible, we ended up with a ten-minute video presentation. This duration was necessary in order to not miss important parts and aspects of our process model. However, for future surveys we recommend introductionary videos with duration of less than 5-8 minutes. The reason for this is that the drop-out rate of participants was quite high at the first two pages (introduction (page 1) and video presentation (page 2)) of the survey. In addition, 8 of 52 participants did not watch the full video but skipped on the next part of the questionnaire after 5 to 8 minutes or earlier, although this part of our survey was mandatory. In order to address this issue, we would recommend defining time spans after which the next part of the survey can be accessed. In our context, at least the duration of the video should have been defined as the minimum amount of time which would have to be spent on page 2 of our survey. However, we deliberately decided not to use this feature, due the risk of losing even more participants if we would disable the “next” button at this early part of the survey. In this context, the definition of mandatory and optional questions needs to be tested and well thought through.

(3) The application of a web-based survey tool like SoSci Survey proved to be applicable to reach larger populations of experts. We recommend this type of survey platforms. Especially in combination with e.g. the interactive click-through presentation and the introductory presentation video of our process model, the application of a web-based survey tool worked well. The introduction of our process model in the form of a webinar-like video presentation did not only allow us

to reach a higher population of experts and participants but also reduced the potential bias which could have been cause by different and varying process model introduction and presentation styles, formats durations and interviewers.

(4) The collection of textual comments in addition to the quantitative likert-scale based statements also showed to be helpful and – as expected – provided us with further insight on our artefact and helped us understand participants answer options.

The main findings are summarized in Table 1:

Table 1: Summary of key learning and recommendations

Title	Description
Importance of participant recruiting process	Definition of a systematic and structured approach of how to approach and remind potential participants is of crucial relevance for getting a satisfactory response rate.
Consideration of timing and duration aspects	The total duration of a web-based survey should not exceed 20-30 minutes. Otherwise, higher drop-out rates and ultimately fewer responses may be the result. If introductionary presentations are part of the survey, the length of these should be kept as short as possible.
Potential of web-based, interactive and multi-media survey instruments	Web-based survey instruments offer the possibility to reach larger groups of participants and allow for conducting survey independent of place and time.
	The provision of the screen-mode presentation to participants allowed for getting additional information after process model introduction and was easy to integrate in the selected survey tool.
	In order to reduce bias caused by different and varying ways of presenting the artefact and required for online presentation of complex models or evaluands, webinar-like videos can be recommended and proved to be applicable in such contexts.
Collection of textual comments	The collection of textual comments for selected statements allowed for gaining further insights into our model from participants' point of view and helped interpreting results.
Mandatory and optional parts of the survey	The division into mandatory and optional survey parts allows for reducing the risk of too long survey durations on the one hand but bears the risk of missing valuable results on the other hand. Optional survey parts should hence only be applied if the respective part is not of basic relevance for the survey's purpose but would e.g. allow for additional interpretation of results.

5 Conclusion

In this paper, we presented an approach to artefact evaluation, which goes beyond the analysis of usefulness of prototypes in the course of case studies and interviews. The web-based survey with interactive and multi-media elements proved to be capable of collecting results from specific expert groups. The evaluation approach developed in order to evaluate our process model from an ex-post perspective provided us with valuable results and showed that there is more to artefact evaluation than just utility of prototypes: based on the three dimensions of syntactical, perceived semantic and perceived pragmatic model quality, we were able to evaluate our artefact. Furthermore, the application of the semi-quantitative web-based survey addressing PSQ and PU items provided us with the possibility to access a high population of experts. The collection of textual statements allowed for gaining further insights and enabled us to derive additional interpretations of survey results. Indeed, our research supports the general observation that the combination of qualitative and quantitative input can lead to new insights and modes of analysis (Venkatesh et al. 2013). However, as our recommendations above indicate there are important elements for consideration. This paper also described the methodological steps of developing a multi-media evaluation approach. These can be adopted by other researchers in similar situations in developing appropriate evaluation approaches in their specific settings. The findings together with some recommendations are presented in section 4 of the paper. Similarly, the recommendations can also be applied in other research setting, where web-based surveys are to be conducted and applied.

Future research should analyse the applicability of semi-quantitative web-based evaluation approaches similar to ours for e.g. different types of artefacts and in additional research domains.

References

Adipat, Boonlit; Zhang, Dongsong; Zhou, Lina (2011): The effects of tree-view based presentation adaptation on mobile web browsing. In *MIS Quarterly* 35 (1), pp. 99–122.

Brandtner, Patrick; Helfert, Markus; Auinger, Andreas; Gaubinger, Kurt (2015): Conducting focus group research in a design science project: Application in developing a process model for the front end of innovation. In *Systems, Signs & Actions* 9 (1), pp. 26–55.

Brown, Julie A.; Serrato, Carl A.; Hugh, Mildred; Kanter, Michael H.; Spritzer, Karen L.; Hays, Ron D. (2016): Effect of a post-paid incentive on response rates to a web-based survey. In *Survey Practice* 9 (1).

Goldkuhl, Göran (2013): Activity Cycles in Design Research: A Pragmatic Conceptualisation of Inter-related Practices. In Markus Helfert, Brian Donnellan (Eds.): Design Science: Perspectives from Europe, vol. 388, pp. 49–60.

Greenlaw, Corey; Brown-Welty, Sharon (2009): A comparison of web-based and paper-based survey methods: testing assumptions of survey mode and response cost. In *Evaluation review* 33 (5), pp. 464–480.

Helfert, Markus; Donnellan, Brian; Ostrowski, Lukasz (2012): The case for design science utility and quality. Evaluation of design science artifact within the sustainable ICT capability maturity framework. In *An International Journal on Information Technology, Action, Communication and Workpractices* 6 (1), pp. 46–66.

Krogstie, John; Nossum, Alexander (2014): A Semiotic Approach for Guiding the Visualizing of Time and Space in Enterprise Models. In Kecheng Liu, Stephen R. Gulliver, Weizi Li, Changrui Yu (Eds.): *Service Science and Knowledge Innovation*, vol. 426. Berlin, Heidelberg, pp. 74–86.

Krogstie, John; Sindre, Guttorm; Jørgensen, Håvard (2006): Process models representing knowledge for action. A revised quality framework. In *Eur J Inf Syst* 15 (1), pp. 91–102.

Liu, Lei; Wu, Yu-Feng; Li, Xiao-jun (2012): Current Practice in Measuring the Quality of Conceptual Models: Challenges and Research. In *International Journal of Plant Engineering and Management* 17 (3).

Maes, Ann; Poels, Geert (2007): Evaluating quality of conceptual modelling scripts based on user perceptions. In *Data & Knowledge Engineering* 63 (3), pp. 701–724.

March, Salvatore T.; Smith, Gerald F. (1995): Design and natural science research on information technology. In *Decision Support Systems* 15 (4), pp. 251–266.

Markus, M. Lynne; Majchrzak, Ann; Gasser, Les (2002): A design theory for systems that support emergent knowledge processes. In *MIS Quarterly*, pp. 179–212.

Niederman, Fred; March, Salvatore T. (2012): Design science and the accumulation of knowledge in the information systems discipline. In *ACM Trans. Manage. Inf. Syst.* 3 (1), pp. 1–15.

Ostrowski, Lukasz; Helfert, Markus (2012): Design Science Evaluation – Example of Experimental Design. In *Journal of Emerging Trends in Computing and Information Sciences* 3 (9), pp. 253–262.

Peffers, Ken; Rothenberger, Marcus; Tuunanen, Tuure; Vaezi, Reza (2012): Design Science Research Evaluation. In Ken Peffers, Marcus Rothenberger, Bill Kuechler (Eds.): *Design Science Research in Information Systems. Advances in Theory and Practice*, vol. 7286: Springer Berlin Heidelberg (Lecture Notes in Computer Science), pp. 398–410.

Peffers, Ken; Tuunanen, Tuure; Rothenberger, Marcus; Chatterjee, Samir (2007): A Design Science Research Methodology for Information Systems Research. In *J. Manage. Inf. Syst* 24 (3), pp. 45–77.

Prat, Nicolas; Comyn-Wattiau, Isabelle; Akoka, Jacky (2014): artifact evaluation in information systems design-science research – a holistic view In *Proceedings of 18th Pacific Asia Conference on Information Systems (PACIS)*, Chengdu, China, p. 17.

Pries-Heje, Jan; Baskerville, Richard; Venable, John (2008): Strategies for Design Science Research Evaluation. In William Golden, Tom Acton, kieran Conboy, Virpi Kristiina Tuunainen (Eds.): Conference Proceedings of 16th European Conference on Information Systems: National University of Ireland, pp. 12.

Reeder, B.; Revere, D.; Olson; Lober, W. B. (2011): Perceived usefulness of a distributed community-based syndromic surveillance system: a pilot qualitative evaluation study. In *BMC research notes* 4, pp. 187

Rittgen, Peter (2010): Quality and perceived usefulness of process models. In *Proceedings of the 2010 ACM Symposium on Applied Computing (ACM)*, pp. 65–72.

Siau, Keng; Rossi, Matti (2011): Evaluation techniques for systems analysis and design modelling methods - a review and comparative analysis. In *Information Systems Journal* 21 (3), pp. 249–268.

Sonnenberg, Christian; Brocke, Jan (2012): Evaluation Patterns for Design Science Research Artefacts. In Markus Helfert, Brian Donnellan (Eds.): *Practical Aspects of Design Science*, vol. 286: Springer Berlin Heidelberg (Communications in computer and information science), pp. 71–83.

Venable, John; Pries-Heje, Jan; Baskerville, Richard (2012): A Comprehensive Framework for Evaluation in Design Science Research. In Ken Peffers, Marcus Rothenberger, Bill Kuechler (Eds.): *Design Science Research in Information Systems. Advances in Theory and Practice*, vol. 7286: Springer Berlin Heidelberg (Lecture Notes in Computer Science), pp. 423–438.

Venkatesh, Viswanath; Brown, Susan A.; Bala, Hillol (2013): Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. In *MIS Quarterly* 37 (1), pp. 21–54.

Venkatesh, Viswanath; Davis, Fred D. (2000): A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. In *Management Science* (46), pp. 186–204.