A Comprehensive Framework for Evaluation in Design Science Research

Conference Paper · May 2012

DOI: 10.1007/978-3-642-29863-9_31

3 authors:

John Robert Venable
Curtin University

Jan Pries-Heje
Roskilde University

Richard Baskerville
Georgia State University

A Comprehensive Framework for Evaluation in Design Science Research[1]

John Venable[1], Jan Pries-Heje[2] and Richard Baskerville[3]

[1] Curtin University, Perth, Western Australia, Australia
j.venable@curtin.edu.au
[2] Roskilde University, Roskilde, Denmark
janph@ruc.dk
[3] Georgia State University, Atlanta, Georgia, USA
baskerville@acm.org

Abstract. Evaluation is a central and essential activity in conducting rigorous Design Science Research (DSR), yet there is surprisingly little guidance about designing the DSR evaluation activity beyond suggesting possible methods that could be used for evaluation. This paper extends the notable exception of the existing framework of Pries-Heje et al [11] to address this problem. The paper proposes an extended DSR evaluation framework together with a DSR evaluation design method that can guide DSR researchers in choosing an appropriate strategy for evaluation of the design artifacts and design theories that form the output from DSR. The extended DSR evaluation framework asks the DSR researcher to consider (as input to the choice of the DSR evaluation strategy) contextual factors of goals, conditions, and constraints on the DSR evaluation, e.g. the type and level of desired rigor, the type of artifact, the need to support formative development of the designed artifacts, the properties of the artifact to be evaluated, and the constraints on resources available, such as time, labor, facilities, expertise, and access to research subjects. The framework and method support matching these in the first instance to one or more DSR evaluation strategies, including the choice of ex ante (prior to artifact construction) versus ex post evaluation (after artifact construction) and naturalistic (e.g., field setting) versus artificial evaluation (e.g., laboratory setting). Based on the recommended evaluation strategy(ies), guidance is provided concerning what methodologies might be appropriate within the chosen strategy(ies).

Keywords: Design Science Research, Research Methodology, Information Systems Evaluation, Evaluation Method, Evaluation Strategy.

---

[1] Venable, John, Pries-Heje, Jan, & Baskerville, Richard. (2012). A Comprehensive Framework for Evaluation in Design Science Research. In K. Peffers, M. Rothenberger & B. Kuechler (Eds.), *Design Science Research in Information Systems. Advances in Theory and Practice* (Vol. 7286, pp. 423-438). Berlin / Heidelberg: Springer.

1 Introduction

There is widespread agreement that evaluation is a central and essential activity in conducting rigorous Design Science Research (DSR). In DSR, evaluation is concerned with examining DSR outputs, including design artifacts [6] and Information Systems (IS) Design Theories [3], [20]. March and Smith [6] identify "build" and "evaluate" as two DSR activities. Hevner et al [5] identify evaluation as "crucial" (p. 82). In their third guideline for Design Science in IS Research, they state that "The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods" (p. 85).

Evaluation provides evidence that a new technology developed in DSR "works" or achieves the purpose for which it was designed. Without evaluation, outcomes of DSR are unsubstantiated assertions that the designed artifacts, if implemented and deployed in practice, will achieve their purpose. Rigorous, scientific research requires evidence. If Design Science Research is to live up to its label as "science", the evaluation must be sufficiently rigorous.

But how should rigorous evaluation be designed and conducted? What strategies and methods should be used for evaluation in a particular DSR project? How can the evaluation be designed to be both effective (rigorous) and efficient (prudently using resources, including time)? What would constitute good guidance for answering these questions?

Unfortunately, there is little guidance in the DSR literature about the choice of strategies and methods for evaluation in DSR. A notable exception is Pries-Heje et al [11], who develop a 2-by-2 framework to guide selection of evaluation strategy(ies) for a DSR project. They identify that evaluation design needs to decide what will be evaluated, when it will be evaluated, and how it will be evaluated. However, beyond providing the framework and an idea of what needs to be designed in the DSR component of research, they provide very little guidance in how a research should or could actually design the DSR evaluation component. This state of affairs in DSR constitutes what we can call an "evaluation gap".

The purpose of this paper is to address this evaluation gap by developing a DSR evaluation framework with clear guidance for how one could design and conduct evaluation within DSR. Making a strong, published evaluation framework available to design science researchers, particularly novice ones, can simplify the research design and reporting. Such guidance would help DSR researchers make decisions about how they can (and perhaps should) conduct the evaluation activities of DSR.

It is important to clarify that the framework developed here is to aid DSR researchers in the design of the evaluation component of their DSR. The framework proposed here is not a framework for evaluating DSR projects as a whole or after the fact. Conducting DSR involves much more than the evaluation of the resulting DSR artifacts and IS Design Theories and such broader evaluation of a whole DSR project is outside the scope of this paper.

This next section of this paper discusses relevant literature on evaluation in DSR to elucidate the "evaluation gap" addressed in this paper. Section 3 describes an extended framework and method developed to address this gap. Section 4 describes the evaluation of the method in use by novice design science researchers. Finally section 5 discusses the findings and presents conclusions.

2 Literature on Evaluation in DSR

This section considers the DSR literature concerning the purposes for evaluation in DSR, characteristics or aspects to be evaluated in DSR evaluation, kinds of artifacts (evaluands) in DSR, design goals to be addressed in the design of a DSR evaluation method, methods proposed for evaluation in DSR, and guidance for designing the evaluation component of DSR.

2.1 Purposes of Evaluation in DSR

As noted above, evaluation is what puts the "Science" in "Design Science". Without evaluation, we only have an unsubstantiated design theory or hypothesis that some developed artifact will be useful for solving some problem or making some improvement. This section identifies and discusses five different purposes for evaluation in the DSR literature.

1. Evaluate an instantiation of a designed artifact to establish its utility and efficacy (or lack thereof) for achieving its stated purpose

March and Smith [6] define evaluation as "the process of determining how well the artifact performs." (p. 254). The central purpose of DSR evaluation then is to rigorously demonstrate the utility of the artifact being evaluated (known as the "evaluand" (Stufflebeam 2000)). DSR design artifacts "are assessed against criteria of value or utility – does it work?" [6]. A key purpose of DSR evaluation then is to determine whether or how well the developed evaluand achieves its purpose.

2. Evaluate the formalized knowledge about a designed artifact's utility for achieving its purpose

Evaluating the design artifact's utility for purpose is closely related to the concepts of IS Design Theories (ISDTs) [3], 18], [20], design principles [7], [10], [12], or technological rules [16], which are formalizations of knowledge about designed artifacts and their utility. When an artifact is evaluated for its utility in achieving its purpose, one is also evaluating a design theory that the design artifact has utility to achieve that purpose. From the point of view of design theory, a second purpose of evaluation in DSR is to confirm or disprove (or enhance) the design theory.

3. Evaluate a designed artifact or formalized knowledge about it in comparison to other designed artifacts' ability to achieve a similar purpose

In addition to the first purpose above, Venable [17] identifies a third purpose – evaluating the artifact "in comparison to other solution technologies" (p. 4). A new artifact should provide greater relative utility than existing artifacts that can be used to achieve the same purpose.

4. Evaluate a designed artifact or formalized knowledge about it for side effects or undesirable consequences of its use

Another purpose that Venable [17] identifies is evaluating an artifact for other (undesirable) impacts in the long run, i.e. for side effects (particularly dangerous ones).

5. Evaluate a designed artifact formatively to identify weaknesses and areas of improvement for an artifact under development

A fifth purpose of evaluation is formative evaluation, in which an artifact still under development is evaluated to determine areas for improvement and refinement. Sein et al [12] use evaluation formatively in early (alpha) Building, Intervention, and Evaluation (BIE) cycles (*cf.* ex ante evaluation in [11]) of their Action Design Research Methodology (ADR). The last BIE cycle in ADR is summative evaluation of a beta version, which is in line with the first purpose for evaluations given above.

Next we turn our attention to what is evaluated.

2.2 Aspects and Characteristics to be Evaluated in DSR

Utility is a complex concept and not the only thing that is evaluated in DSR. Utility may depend on a number of characteristics of the artifact or desired outcomes of the use of the artifact. Care must be taken to consider how utility for achieving the artifact's purpose(s) can be assessed, what characteristics to evaluate or measure. Each evaluation is quite specific to the artifact, its purpose(s), and the purpose(s) of the evaluation.

Nonetheless, it is useful to consider what kinds of qualities for evaluation are discussed in the literature. As noted earlier, Hevner et al [5] identify utility, quality, and efficacy as attributes to be evaluated. Hevner et al [5] further state that "artifacts can be evaluated in terms of functionality, completeness, consistency, accuracy, performance, reliability, usability, fit with the organization, and other relevant quality attributes" (p. 85). They later identify "style" as an aspect of an artifact that should be evaluated.

Checkland and Scholes [1] proposed five properties ("the 5 E's") by which to judge the quality of an evaluand: Efficiency, effectiveness, efficacy, ethicality, and elegance. Effectiveness and efficacy are sometimes confused. Effectiveness is the degree to which the artifact meets its higher level purpose or goal and achieve its desired benefit in practice. Efficacy is the degree to which the artifact produces its desired effect considered narrowly, without addressing situational concerns.

All of these properties of the artifact in some way contribute to the utility of the developed artifact and act as criteria that are candidates for evaluation in determining the overall utility.

2.3 Kinds of Evaluands in DSR

Next we consider the different kinds of evaluands. Based on the literature, we can identify two different classifications of artifacts.

First, we can distinguish product artifacts from process artifacts [3], [18]. Product artifacts are technologies such as tools, diagrams, software, etc. that people use to accomplish some task. Process artifacts are methods, procedures, etc. that guide someone or tell them what to do to accomplish some task.

Second, we can distinguish between technical artifacts and socio-technical artifacts. Some artifacts are in some sense "purely" (or nearly purely) technical, in that they do not require human use once instantiated. Socio-technical artifacts are ones with which humans must interact to provide their utility.

Relating the technical vs socio-technical distinction to the product vs process distinction, product artifacts may be either (purely) technical or socio-technical, while process artifacts are always socio-technical, which will have implications for their evaluation.

2.4 Goals of Evaluation Design in DSR

Next we consider what goals there are for the design of the evaluation itself. There are (at least) three possibly competing goals in designing the evaluation component of DSR.

- Rigor: Research, including DSR, should be rigorous. Rigor in DSR has two senses. The first is in establishing that it is the artifact (instantiation) that causes an observed improvement (and only the artifact, not some confounding independent variable or circumstance), i.e. its efficacy. The second sense of rigor in DSR is in establishing that the artifact (instantiation) works in a real situation (despite organisational complications, unanticipated human behavioral responses, etc.), i.e. its effectiveness.
- Efficiency: A DSR evaluation should work within resource constraints (e.g. money, equipment, and people's time) or even minimize their consumption.
- Ethics: Research, including DSR, should not unnecessarily put animals, people, organizations, or the public at risk during or after evaluation, e.g. for safety critical systems and technologies. Venable [19] discusses some ethical issues in DSR.

The 5 E's [1] are also relevant to the design of the evaluation part of a DSR project. Each of the above goals corresponds to one of the 5 E's. Only Elegance is missing, although presumably an elegant evaluation would be preferable to an inelegant one. Importantly these goals conflict and DSR evaluation must balance these goals.

2.5 Evaluation Methods in DSR

Next we consider what methods there are for evaluation (from which a Design Science researcher might choose).

Different DSR authors have identified a number of methods that can be used for evaluation in DSR. Hevner et al [5] summarize five classes of evaluation methods with 12 specific methods in those classes. (1) Observational methods include case study and field study. (2) Analytical methods include static analysis, architecture analysis, optimization, and dynamic analysis. (3) Experimental methods include controlled experiment and simulation. (4) Testing methods include functional (black box) testing and structural (white box) testing. (5) Descriptive methods include informed argument and scenarios. They provide no guidance on method selection or evaluation design.

Vaishnavi and Kuechler [15] allow for both quantitative and qualitative methods and describe the use of a non-empirical analysis. They do not provide guidance for selecting between methods or designing the evaluation part of DSR.

Peffers et al [9] divide what others call evaluation into two activities, demonstration and evaluation. Demonstration is like a light-weight evaluation to demonstrate that the artifact feasibly works to "solve one or more instances of the problem", i.e. to achieve its purpose in at least one context (*cf.* ex ante evaluation in [11]). Evaluation proper is more formal and extensive, and takes a fairly positivistic stance that the activity should evaluate "how well the artifact supports a solution to the problem" (p. 56). Methods for evaluation identified include the collection of objective quantitative performance measures such as budgets or items produced, the results of satisfaction surveys, client feedback" (p. 56), or the use of simulations or logical proofs, but they provide no guidance for choosing between methods.

Nunamaker et al [8] identified a number of methods for evaluation or what they termed experimentation. These included computer and lab simulations, field experiments, and lab experiments. Additionally, they identified several methods of observation, including case studies, survey studies, and field studies, although they did not see these as evaluation methods. Moreover, they did not provide much guidance in choosing among these evaluation methods, except to say that the evaluation method must be matched to the designed artifact and the evaluation metrics to be used.

The activities that Nunamaker et al [8] called experimentation and observation, Venable [17] instead respectively called artificial evaluation and naturalistic evaluation, explicitly recognizing the evaluative nature of the observation activity. Artificial evaluation includes laboratory experiments, field experiments, simulations, criteria-based analysis, theoretical arguments, and mathematical proofs. The dominance of the scientific/rational paradigm brings to artificial DSR evaluation the benefits of stronger scientific reliability in the form of better repeatability and falsifiability [4].

Naturalistic evaluation explores the performance of a solution technology in its real environment i.e., within the organization. By performing evaluation in a real environment (real people, real systems, and real settings [14]), naturalistic evaluation embraces all of the complexities of human practice in real organizations. Naturalistic evaluation is always empirical and may be interpretive, positivist, and/or critical. Naturalistic evaluation methods include case studies, field studies, surveys, ethnography, phenomenology, hermeneutic methods, and action research. The dominance of the naturalistic paradigm brings to naturalistic DSR evaluation the benefits of stronger internal validity [4].

Artificial and naturalistic evaluation each have their strengths and weaknesses. To the extent that naturalistic evaluation is affected by confounding variables or misinterpretation, evaluation results may not be precise or even truthful about an artifact's utility or efficacy in real use. On the other hand, artificial evaluation involves abstraction from the natural setting and is necessarily "unreal" according to one or more of Sun and Kantor's [14] three realities (unreal users, unreal systems, or unreal problems). To the extent that an artificial evaluation setting is unreal, evaluation results may not be applicable to real use. In contrast, naturalistic evaluation offers more critical face validity. Evaluation in a naturalistic setting is "the real 'proof of the pudding'" [17, p. 5].

Further, Venable noted that more than one method could be used, mixing artificial and naturalistic evaluation as well as positivist and interpretive evaluation methods, leading to a pluralist view of science, where each has its strengths in contributing to a robust evaluation depending on the circumstance. Nonetheless, Venable [17] provided little or no guidance about selecting among methods and designing an evaluation strategy.

In summary, the DSR literature identifies a fairly large number and variety of evaluation methods, but gives little advice as to choice among methods, i.e. how to design an evaluation strategy for a particular DSR project.

2.6 Guidance for Designing Evaluations in DSR

While the DSR literature provides almost no guidance on how to design the evaluation component of DSR research, there is one notable exception: the paper by Pries-Heje et al [11], which proposes a 2-by-2 framework of strategies for evaluation in DSR (see figure 1 below) and provides some guidance for considerations about how to choose among them. Their framework combines one dimension contrasting artificial vs naturalistic evaluation [17], as discussed in section 2.5, with a second dimension contrasting ex ante and ex post evaluation. Ex post evaluation is evaluation of an instantiated artifact (i.e. an instantiation) and ex ante evaluation is evaluation of an uninstantiated artifact, such as a design or model. This distinction is similar to the later distinction in ADR concerning evaluation of alpha versions of an artifact for formative purposes vs evaluation of beta versions of an artifact for summative purposes [12]. The paper also takes into account that what is being evaluated – the design artifact - can either be a process, a product or both (as discussed in section 2.3).

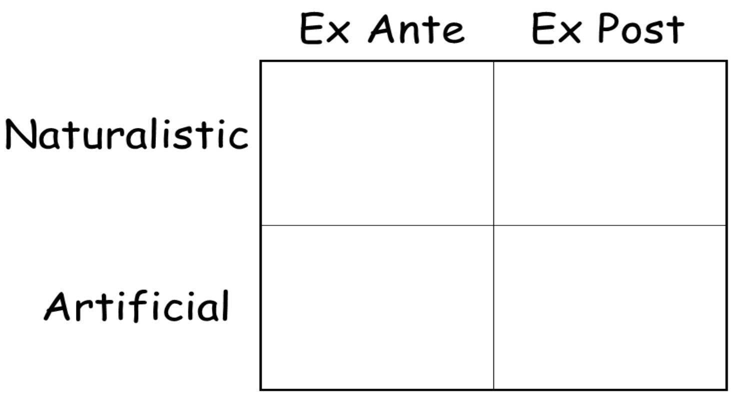

Fig. 1. A Strategic DSR Evaluation Framework (adapted from [11])

Some key points that Pries-Heje et al [11] make concerning the design of evaluation in DSR are that:

1. The distinctions of ex ante vs ex post and artificial vs naturalistic evaluation surface a variety of ways in which evaluation might be conducted.
2. Ex ante evaluation is possible and building an instantiation of an artifact may not be needed (at least initially).
3. Artifact evaluation in artificial settings could include imaginary or simulated settings.
4. Naturalistic evaluation can be designed by choosing from among multiple realities and multiple levels of granularity for measurements or metrics.
5. Multiple evaluations, combining multiple evaluation strategies, may be useful.
6. The specific evaluation criteria, measurements, or metrics depend on the type of artifact (product or process) and intended goals or improvements.

While the above suggestions to guide the research design of evaluation in DSR are useful, we believe they are incomplete and less useful than they might be. There is no guidance for considering how the different purposes, evaluation design goals, available resources, etc. can or should be considered when choosing a DSR evaluation strategy or strategies. Moreover, they provide no guidance about how to select evaluation methods. These difficulties are addressed in the next section.

3 A Comprehensive Framework and Method for Designing Evaluation in Design Science Research

In this section, we develop an extended and comprehensive framework and method for designing the evaluation method(s) used in a particular DSR project.

The comprehensive DSR framework and method needs to provide support for deriving the design of the DSR project's evaluation method from an understanding of the DSR project context, including the desired evaluation purpose, goals, and practical constraints. The framework should help to identify a particular DSR evaluation strategy (or combination of strategies) that is appropriate and also to support decision making about what particular evaluation method(s) are appropriate (possibly best or optimal) to achieve those strategies.

The method and framework we have developed in this paper extends the framework described in [11]. The extensions are in three parts: (1) a framework extension to map evaluation purpose, goals, and artifact type as contextual aspects that set the criteria for the evaluation design to a potential evaluation strategy or strategies (see figure 2 and section 3.1), (2) an extended framework to map a chosen evaluation strategy or strategies to candidate evaluation methods (see figure 3 and section 3.2), and (3) a process or method to use the two extended frameworks (see section 3.3).

3.1 First Extension: A DSR Evaluation Strategy Selection Framework

The first extension relates or maps various aspects of the context of the evaluation in the DSR project to the framework by Pries-Heje et al [11], as shown in figure 2 (A DSR Evaluation Strategy Selection Framework). Relevant aspects of the context of the DSR evaluation serve as the starting point and input to the design of the DSR evaluation. Relevant contextual aspects include (1) the different purposes of evaluation in DSR, (2) the characteristics of the evaluand to be evaluated, (3) the type of evaluand to be evaluated, and (4) the specific goals that must be balanced in the design of the evaluation part(s) of a DSR project. These four contextual aspects were discussed in sections 2.1 through 2.4 respectively.

In figure 2, the above four contextual aspects are combined into criteria that should be considered as input to the DSR evaluation design. These criteria include the following and are mapped to ex ante vs ex post and artificial vs naturalistic evaluation as shown in the white areas of figure 2,

- The extent to which cost and time resource limitations constrain the evaluation or the whole research project
- Whether (or not) early, formative evaluation is desirable and feasible
- The extent to which the artifact being designed has to please heterogeneous groups of stakeholders or if there is likely to be conflict, which will complicate evaluation
- Whether the system is purely technical in nature or socio-technical in nature, with the consequent difficulties of the latter (cf. artifact focus as either technical, organizational, or strategic [2]
- How important strong rigor concerning effectiveness in real working situations is
- How important strong rigor concerning whether benefits are specifically due to the designed artifact, rather than some other potential cause (or confounding variable), is
- Whether or not access to a site for naturalistic evaluation is available or can be obtained
- Whether the level of risk for evaluation participants is acceptable or needs to be reduced

To use the framework, a design science researcher begins with an understanding of the context of the DSR evaluation, maps that understanding to the criteria in figure 2, and selects an evaluation strategy or combination of strategies based on which rows, columns and cells in figure 2 are most relevant.

DSR Evaluation Strategy Selection Framework		**Ex Ante**	**Ex Post**
		•Formative •Lower build cost •Faster •Evaluate design, partial prototype, or full prototype •Less risk to participants (during evaluation) •Higher risk of false positive	•Summative •Higher build cost •Slower •Evaluate instantiation •Higher risk to participants (during evaluation) •Lower risk of false positive
Naturalistic	•Many diverse stakeholders •Substantial conflict •Socio-technical artifacts •Higher cost •Longer time - slower •Organizational access needed •Artifact effectiveness evaluation •Desired Rigor: "Proof of the Pudding" •Higher risk to participants •Lower risk of false positive – safety critical systems	•Real users, real problem, and somewhat unreal system •Low-medium cost •Medium speed •Low risk to participants •Higher risk of false positive	•Real users, real problem, and real system •Highest Cost •Highest risk to participants •Best evaluation of effectiveness •Identification of side effects •Lowest risk of false positive – safety critical systems
Artificial	•Few similar stakeholders •Little or no conflict •Purely technical artifacts •Lower cost •Less time - faster •Desired Rigor: Control of Variables •Artifact efficacy evaluation •Less risk during evaluation •Higher risk of false positive	•Unreal Users, Problem, and/or System •Lowest Cost •Fastest •Lowest risk to participants •Highest risk of false positive re. effectiveness	•Real system, unreal problem and possibly unreal users •Medium-high cost •Medium speed •Low-medium risk to participants

Fig. 2. A DSR Evaluation Strategy Selection Framework

In using figure 2 to formulate a DSR evaluation strategy or strategies, it is important to prioritize these different criteria, as they are likely to conflict. For example, obtaining the rigor of naturalistic evaluation my conflict with reducing risk to evaluation participants and the need to reduce costs. If cost and risk reduction override (or preclude) rigorous evaluation of effectiveness in real settings, then an artificial evaluation strategy may be chosen as more appropriate.

In formulating an evaluation strategy, figure 2 can advise the DSR researcher in the choice. Identifying relevant, higher priority criteria in the white and blue cells supports identifying an appropriate quadrant or quadrants, i.e. the relevant blue cell(s) in figure 2. Note that picking a single box may not be the best strategy; rather, a hybrid strategy (more than one quadrant) can be used to resolve conflicting goals.

3.2 Second Extension: A DSR Evaluation Method Selection Framework

The second extension is to relate the different evaluation strategies in the framework by Pries-Heje et al [11] to different extant evaluation methods, which were also discussed in section 2.5. This extension is expressed as a mapping of DSR evaluation strategies to relevant evaluation methods (see figure 3). By combining these two figures, the extended framework provides a bridge between the contextual factors relevant to the DSR evaluation and appropriate means (methods) to evaluate the DSR artifacts.

DSR Evaluation Method Selection Framework	Ex Ante	Ex Post
Naturalistic	•Action Research •Focus Group	•Action Research •Case Study •Focus Group •Participant Observation •Ethnography •Phenomenology •Survey (qualitative or quantitative)
Artificial	•Mathematical or Logical Proof •Criteria-Based Evaluation •Lab Experiment •Computer Simulation	•Mathematical or Logical Proof •Lab Experiment •Role Playing Simulation •Computer Simulation •Field Experiment

Fig. 3. A DSR Evaluation Method Selection Framework

Having decided the high level strategy to be used for evaluation (i.e. which of the quadrants in Figure 2 will be used for the evaluation), then the particular evaluation research method(s) need to be chosen and the evaluation designed in detail. Figure 3 gives a mapping of different possible DSR evaluation research methods map into each quadrant of the framework in Figures 1 and 2. This mapping may omit some potential evaluation methods and other evaluation methods may be developed or adopted for DSR.

Depending on which quadrant(s) were chosen as the DSR evaluation strategy (using figure 2), figure 3 suggests possible evaluation methods that fit the chosen evaluation strategy. The specific choice of evaluation method or methods requires substantial knowledge of the method(s). If the DSR researcher is unfamiliar with the possible methods, he or she will need to learn about them. Further characteristics of the evaluation method will need to be assessed against the specific goals and other contextual issues of the specific DSR project. Detailed advice on which method or methods to select to fit a particular DSR evaluation strategy is therefore beyond the scope and available space of this paper.

3.3 A Four-Step Method for DSR Evaluation Research Design

The third extension is a four-step DSR evaluation research design method that relies on the extended framework as shown in figures 2 and 3.

The development of the extended framework as elaborated in figures 2 and 3, together with our collective experience conducting and supervising DSR projects, enables us to deduce and design a four-step method (or process) for designing the evaluation component(s) of a DSR project. In general, these are to (1) analyze the requirements for the evaluation to be designed, (2) map the requirements to one or more of the dimensions and quadrants in the framework using figure 2, (3) select an appropriate evaluation method or methods that align with the chosen strategy quadrant(s) using figure 3, and (4) design the evaluation in more detail.

1. Analyze the context of the evaluation – the evaluation requirements
   As a first step, we need to identify, analyze, and priorities all of the requirements or goals for the evaluation portion of the DSR project.
   a. Determine what the evaluands are/will be. Will they be concepts, models, methods, instantiations, and/or design theories?
   b. Determine the nature of the artifact(s)/evaluand(s). Is (are) the artifact(s) to be produced a product, process, or both? Is (are) the artifact(s) to be produced purely technical or socio-technical? Will it (they) be safety critical or not?
   c. Determine what properties you will/need to evaluate. Which of these (and/or other aspects) will you evaluate? Do you need to evaluate utility/effectiveness, efficiency, efficacy, ethicality, or some other quality aspect (and which aspects)?
   d. Determine the goal/purpose of the evaluation. Will you evaluate single/main artifact against goals? Do you need to compare the developed artifact against with other, extant artifacts? Do you need to evaluate the developed artifact(s) for side effects or undesired consequences (especially if safety critical)?

e. Identify and analyze the constraints in the research environment. What resources are available – time, people, budget, research site, etc.? What resources are in short supply and must be used sparingly?

f. Consider the required rigor of the evaluation. How rigorous must the evaluation be? Can it be just a preliminary evaluation or is detailed and rigorous evaluation required? Can some parts of the evaluation be done following the conclusion of the project?

g. Prioritize the above contextual factors to determine which aspects are essential, more important, less important, nice to have, and irrelevant. This will help in addressing conflicts between different evaluation design goals.

2. Match the needed contextual factors (goals, artifact properties, etc.) of the evaluation (from step 1) to the criteria in figure 2 ("DSR Evaluation Strategy Selection Framework") , looking at the criteria in both white portions relating to a single dimension and the blue areas relating to a single quadrant. The criteria statement that match the contextual features of your DSR project will determine which quadrant(s) applies(y) most or are most needed. It may well be that more than one quadrant applies, indicating the need for a hybrid methods evaluation design.
3. Select appropriate evaluation method(s) from those listed in the selected, corresponding quadrant(s) in figure 3 ("DSR Evaluation Method Selection Framework"). If more than one box is indicated, selecting a method present in more than one box may be helpful. The resulting selection of evaluation methods and together with the strategy(ies) (quadrant(s)) constitutes a high level design for the evaluation research.
4. Design the DSR evaluation in detail. Ex ante evaluation will precede ex post evaluation, but more than one evaluation may be performed and more than one method used, in which case the order of their use and how the different evaluations will fit together must be decided. Also, the specific detailed evaluations must be designed, e.g. design of surveys or experiments. This generally will follow the extant research methods literature.

4 Evaluation of the framework

When writing about evaluation it is obvious that the framework derived needs to be evaluated. We should take our own medicine so to say. To some extent we have. For three years, the authors have taught various versions of the evaluation framework as it has evolved to a variety of students and scholars carrying out design science research at our and other universities. They have been taught the four steps presented above as well as different evaluation methods. In particular, at Roskilde University, they have been asked to apply the framework in real DSR projects with an average size between 1 and 2 man years (6 people, 3 months full time, is a typical project).

One example from Roskilde University was a group that redesigned a bike lane to make people behave better when biking, i.e. less rude to other people biking and people walking. They designed with Ockham's Razor of simplicity in mind. They used the theory of planned behaviour to inform their design. The group decided that their redesign should be evaluated with a real user (biker) focusing on real problems, i.e. naturalistically. However, access was a problem since it would not be possible to fully implement the real solution without obtaining a lot of permissions and red tape from the Ministry and the Municipality, suggesting an ex ante naturalistic evaluation instead, based on figure 2 (DSR Evaluation Strategy Selection Framework). Thus they instead chose ex ante evaluation and used a Focus Group for evaluation as suggested by figure 3 (DSR Evaluation Method Selection Framework) and deferred instantiation and ex post evaluation to another project with sufficient access and other resources.

Another example from Roskilde University was a group that designed and constructed a digital collaborative workspace. Here material from an existing but completed project was used to evaluate the project. Material included among other things the requirements specification, the project plan, and some Scrum artifacts on tasks. As a full ex post naturalistic evaluation would be too time consuming and resource intensive for the project group, the team chose an ex post artificial evaluation strategy (shown as appropriate in figure 2) and the project outcome was evaluated using a kind of Computer Simulation of the digital workspace (as suggested by figure 3).

The third example is called ChickenNet. Here the group investigated how the use of a digital learning game can give elderly people the ability to achieve the necessary skills for using the internet. The project was inspired by another project called BrainLounge, the purpose of which is to help elderly to exercise their brain through interactive games. The group used an iterative design process, collecting expert and user feedback after each iteration, i.e. it focussed on formative, ex ante evaluation as suggested by figure 2. The product at the end was a mock-up that illustrated the final game design. Here a combination of naturalistic evaluation (real users) and artificial evaluation (experts) was used. The first iteration was ex ante and then the following iterations moved towards ex post, ending with a mock-up. Again figure 3 turned out to be useful in choosing evaluation method

Overall, the result of the evaluation of our evaluation framework is quite positive. Hundreds of students and scholars (around 500 in total) have been able to use the framework, have made decisions on how to evaluate their design artifact, and have carried out the evaluation in accordance with the comprehensive framework as presented in this paper. In most cases, they chose appropriate evaluation strategies and methods.

Thus far our own evaluation has been naturalistic, ex ante, as the methodology has not stabilized until this writing (indeed it may further evolve). Given the lack of other guidance, the risk to participant users is quite low. As a sociotechnical artifact, a naturalistic evaluation seems natural. During evaluation, we have observed our students, sought more general feedback and listened for suggestions for improvement (as well as deducing our own ideas for improvement based on user reactions and problems experienced). A more formal, rigorous evaluation seeking clear ratings as well as open comments about different aspects of the framework and method and their goals will be sought in the next round of usage. As the risk remains fairly low and the artifact is a socio-technical one, a naturalistic, ex post evaluation is suggested for such rigorous evaluation, perhaps using surveys or focus groups of the method users.

5 Conclusion

Evaluation is a very significant issue in IS Design Science Research, yet there is little guidance concerning how to choose and design an appropriate evaluation strategy.

To address the above need, we have developed and presented three enhancements to the existing DSR Evaluation Strategy Framework proposed by Pries-Heje et al [11], which are based on an analysis and synthesis of works on DSR as presented in section 2. The first part of the extended framework (figure 2) maps aspects of the context of a DSR evaluation, such as resources, purpose, goals, and priorities, to the two dimensions and four quadrants of the Pries-Heje et al [11] DSR Evaluation Strategy Framework. The second part (figure 3) maps the quadrants (or the selected relevant DSR evaluation strategy or strategies) to available and relevant research methods that could be chosen to conduct the evaluation or multiple evaluation episodes. We have further developed a detailed four-step method for the design of the evaluation components in a DSR project. This new framework and method should assist DSR researchers, particularly those new to the field, to improve the efficiency and effectiveness of their DSR evaluation activities.

The primary aim of the enhanced framework and method is to guide the Design Science researchers who may need assistance in deciding how to design the evaluation component of their DSR projects. The framework could also be used by reviewers of DSR publications or research proposals in evaluating research design choices, but that is not our intent.

We have tried out and evaluated the extended framework and method in numerous design research projects, including our own and student projects. Nonetheless, further research is needed to gain more experience using the comprehensive DSR evaluation framework and the DSR evaluation design method, further evaluate their utility, and further develop and improve the method, especially as new DSR evaluation methods are developed.

References

1. Checkland, P., Scholes, J.: Soft Systems Methodology in Practice. J. Wiley, Chichester (1990).
2. Cleven, A., Gubler, P., Hüner, K.: 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 (DESRIST 2009). ACM Press, Malvern, PA, USA (2009)
3. Gregor, S., Jones, D.: The Anatomy of a Design Theory. Journal of the Association for Information Systems, vol. 8, pp. 312-335 (2007)
4. Gummesson, E.: Qualitative Methods in Management Research. Studentlitterature, Chartwell-Bratt, Lund, Sweden (1988)
5. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science In Information Systems Research., MIS Quarterly, vol. 28, pp. 75-105 (2004)

6. March, S.T., Smith, G.F.: Design and natural science research on information technology. Decision Support Systems, vol. 15, pp. 251-266 (1995)
7. Markus, M.L., Majchrzak, A., Gasser, L.“ A design theory for systems that support emergent knowledge processes. MIS Quarterly, vol. 26, pp. 179-212 (2002)
8. Nunamaker, J.F., Chen, M., Purdin, T.D.M.: Systems Development in Information Systems Research. Journal of Management Information Systems, vol. 7, pp. 89-106 (1990/1991).
9. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.“ A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, vol. 24 (2008).
10. Pries-Heje, J., Baskerville, R.: The Design Theory Nexus. MIS Quarterly, vol. 32, pp 731-755 (2008)
11. Pries-Heje, J., Baskerville, R., Venable, J.R.: Strategies for Design Science Research Evaluation. In: Proceedigns of the 16th European Conference on Information Systems (ECIS 2008), Galway, Ireland (2008)
12. Sein, M.K., Henfridsson, O., Purao, S., Rossi, M., Lindgren, R.: Action Design Research. MIS Quarterly, vol. 35, pp. 37-56 (2011)
13. Stufflebeam, D.L.: The Methodology of Metaevaluation as Reflected in Metaevaluations by the Western Michigan University Evaluation Center. pp. 95-125 (2000)
14. Sun, Y., Kantor, P.B.: Cross-Evaluation: A new model for information system evaluation. Journal of the American Society for Information Science and Technology, vol. 57, pp. 614-628 (2006)
15. Vaishnavi, V., Kuechler, W.: Design Research in Information Systems. AISWorld (http://desrist.org/design-research-in-information-systems/ (accessed 3 March 2012) (2004)
16. van Aken, J.E.: Management Research Based on the Paradigm of the Design Sciences: The Quest for Field-Tested and Grounded Technological Rules. Journal of Management Studies, vol. 41, pp 219-246 (2004)
17. Venable, J.R..: A Framework for Design Science Research Activities. In: Proceedings of the 2006 Information Resource Management Association Conference Washington, DC, USA (2006)
18. Venable, J.R.: The Role of Theory and Theorising in Design Science Research. In: Proceedings of the 1st International Conference on Design Science Research in Information Systems and Technology (DESRIST 2006), A.R. Hevner and S. Chatterjee (eds.), Claremont, CA, USA (2006)
19. Venable, J.R.: Identifying and Addressing Stakeholder Interests in Design Science Research: An Analysis Using Critical Systems Heuristics, In: IFIP WG 8.2 Working Conference on The role of IS in leveraging the intelligence and creativity of SME’s (CreativeSME), Guimarães, Portugal, Springer, Heidelberg, pp. 93-112 (2009)
20. Walls, J.G., Widmeyer, G.R., El Sawy, O.A.: Building an information system design theory for vigilant EIS. Information Systems Research, vol. 3, pp. 36-59 (1992)