2008

Strategies for Design Science Research Evaluation

Jan Pries-Heje
Roskilde University, janph@ruc.dk

Richard Baskerville
baskerville@acm.org

John R. Venable
School of Information Systems Curtin University of Technology Perth, Western Australia, j.venable@curtin.edu.au

STRATEGIES FOR DESIGN SCIENCE RESEARCH EVALUATION

Pries-Heje, Jan, Roskilde University, Dept. of Communication, Business & IT, Building 42.2, Universitetsvej 1, 4000 Roskilde, Denmark, janph@ruc.dk

Baskerville, Richard, Georgia State University, CIS Dept., 35 Broad Street NW, PO Box 4015, Atlanta GA 30302, USA, baskerville@acm.org

Venable, John, Curtin University of Technology, School of Information Systems, GPO Box U1987, Perth WA 6845, Australia, j.venable@curtin.edu.au

Abstract

Seminal works in the application of design science research (DSR) in IS emphasize the importance of evaluation. However, discussion of evaluation activities and methods is limited and typically assumes an ex post perspective, in which evaluation occurs after the construction of an IS artifact. Such perspectives can assume that the evaluation is an empirical process and its methods can be selected in the same way as empirical research methods. In this paper, we analyze a broader range of evaluation strategies, which includes ex ante (prior to artifact construction) evaluation. This broader view is developed as a strategic DSR evaluation framework, which expands evaluation choices for IS DSR researchers, and also adds emphasis to strategies for evaluating design processes in addition to design products, using well-known quality criteria as an important asset. The framework encompasses both ex ante and ex post orientations as well as naturalistic settings (e.g., case studies) and artificial settings (e.g., lab experiments) for DSR evaluation. The framework proposed offers a strategic view of DSR evaluation that is useful in analyzing published studies, and also in surfacing the evaluation opportunities that present themselves to IS DSR researchers.

Keywords: Design Science Research, Research Methodology, Information Systems Evaluation, Research Criteria.

1 INTRODUCTION

The scientific view of design arises from the concepts found in Simon's seminal *The Sciences of the Artificial* (1996). Simon's design research involves three fundamental aspects: an imperative or prescriptive logic, a search for alternatives, and the evaluation of design. This paper is concerned with the third of these fundamental aspects, i.e. evaluation, which Simon (1996) more-or-less left open for future development.

Evaluation in Design Science Research (DSR) is concerned with evaluation of design science outputs, including theory and artefacts. Walls et al (1992) proposed IS Design Theories at the output of DSR. March and Smith (1995) framed the outputs of DSR as artefacts: "constructs, models, methods, and instantiations" (p. 256). March and Smith (1995) also identify evaluation as one of two DSR activities (together with "build"). Hevner et al (2004) identify evaluation as "crucial" (p. 82) and require researchers to demonstrate the utility, quality, and efficacy of a design artefact using rigorous evaluation methods. Vaishnavi & Kuechler, (2004) state that designed artefacts must be analyzed as to their use and performance as possible explanations for changes (and hopefully improvements) in the behaviour of systems, people, and organizations. However, none of this literature provides much guidance about the choice of strategies and methods for evaluation in DSR. This paper addresses that gap by developing a framework for choosing among evaluation strategies and methods.

2 EVALUATION IN THE IS RESEARCH LITERATURE

Evaluation has been a topic both in general IS Research and in DSR. In the general IS literature, evaluation is generally regarded from one of two perspectives. In the *ex ante* perspective, candidate systems or technologies are evaluated before they are chosen and acquired or implemented. In the *ex post* perspective, a chosen system or technology is evaluated after it is acquired or implemented (Klecun & Cornford, 2005). We consider each of these two perspectives in turn before turning to evaluation in IS DSR.

2.1 Ex Ante Evaluation

Ex ante evaluation is well developed for the purpose of deciding whether or not to acquire or develop a technology, or for the purpose of deciding which of several competing technologies should be acquired or adopted. In its simplest form, ex ante evaluation operates as a cost benefit analysis. When regarded from the perspective of design research, ex ante evaluation provides models for theoretically evaluating a design without actually implementing the material system or technology. In other words, the artefact is evaluated on the basis of its design specifications alone.

Bannister and Remenyi (2000) take an ex ante perspective and categorize two aspects of evaluation methodologies, which we will represent along two dimensions. The first dimension offers three categories of basic approaches. *Fundamental measures* are metrics that capture characteristics of the technology investment as a single measure. These approaches include capital budgeting, return on investment, user satisfaction ratings, etc. The second category is *composite approaches* that combine several fundamental measures and produce a more complex or comprehensive representation of the value of the investment. Examples include information economics, portfolios, balanced scorecard, etc. The third category is *meta-approaches* which use the context as the basis for selecting the optimum set of measures.

The second dimension characterizes the two different ways in which the approaches can be applied. The first way is *positivist/reductionist*, in which the metrics ultimately determine the decision. The second way is *hermeneutic*, in which decision-makers operate according to their understanding and interpretation of the metrics. Hermeneutics applications recognize instinct and intuition, among other influences, that affect the perception of value held by the decision-maker. Table 1 illustrates the framework.

Table 1. Categories of ex ante evaluation methods (adapted from Bannister & Remenyi, 2000)

		Basic Approach		
		Fundamental	Composite	Meta-
Application	Positivist/Reductionist			
	Hermeneutic			

The ex ante perspective is dominated by the economic concerns of whether the system or technology will be worth its costs. In its classical economics form it operates with a theory of choice. Choice theories distinguish absolute choice from comparative choice. An absolute choice assumes the decision-maker will evaluate each option separately and choose the option assigned the highest value. Comparative choice is a process in which options are compared one against another and the most attractive option surfaces from among its specified alternatives. Sophisticated economic models for both forms of choice theories will often build their complexity on utility theory and will account for such elements in the decision process as decision-maker risk aversion (Shafir et al. 1993).

For information systems and technologies, the ex ante perspective is more complicated in evaluating public sector systems in which the profit motive is less relevant. Economic metrics are often complicated (although not impossible) by measures for human life and well-being (Scarbrough, 1998).

As a result, frameworks for ex ante evaluation of public sector systems sometimes bridge across to ex post evaluation through participatory measures and social perspectives (Irani et al. 2005). For example, Keast and Waterhouse (2006) propose an evaluation approach, in which the ex ante evaluators are formal roles that gradually relinquish authority to ex post participatory self-evaluators.

2.2 Ex post Evaluation

In their work on the evaluation of online personalization systems, Yang and Padmanabhan (2005) categorize ex post evaluation methodologies along two rather different dimensions. The setting dimension distinguishes real settings from settings that are not real. There is a second dimension that distinguishes different methods for the computation of quality measures. Automatic computation is developed directly from fundamental data, and is distinguished from quality measures that have a basis in the opinions of human subjects. Table 2 illustrates the framework, which combines the two dimensions into four combinations/categories.

Table 2. Categories of ex post evaluation methods (adapted from Yang & Padmanabhan, 2005)

		Setting	
		Real Setting	Abstract Setting
Computation of Quality Measures	Automatic	1. Experimental designs	3. Historic data experiments
	Human Subjects	2. User opinion studies	4. Opinions analysis of historical data

This ex post framework provides examples of each of the four categories of evaluation approach. In the first category, classical X-O experimental designs illustrate real situations in which quantitative evaluation data can be collected and calculated by observing actual use. The collection of subjective opinions of the users of actual technologies and systems represents a second category.

Yang and Padmanabhan (2005) explain the alternative to real settings from a perspective of data mining into usage records. For example, in the third category in Table 2, a web personalization technology can be tested by feeding it historical web browsing data. In the fourth category, such abstract test designs can be extended to include comparisons with the analysis or opinions of human subjects. Perhaps the most well-know example of such work is the Delone and McLean's (1992) analysis of research results based on a concept like "success" representing the dependent variable. Their work created an abstract model of the success construct that emerged from the prior scientific history of system performance evaluations.

Contemporary approaches to ex post evaluation in information systems build from Symons' (1991) critical adaptation of the "context, content and process" (CCP) model developed for organizational change evaluation. Examples of variations include interpretive (Stockdale & Standing, 2006) and critical (Klecun & Cornford, 2005) evaluation approaches. The context construct represents the requirement that any evaluation needs to be contextualized according to the needs of the stakeholders in the evaluation. The content construct represents the decisions about what measures or metrics should comprise the evaluation. The process construct represents decisions about the instrumentation and methodologies selected for performing the evaluation. These three constructs capture two of the major themes in the information systems evaluation literature: alternative measurement instruments/constructs such as user satisfaction and system use; and alternative approaches to the evaluation process (Stockdale & Standing, 2006).

Contemporary work from the information science field is also oriented toward ex post evaluation, but from a perspective of evaluating the knowledge or information content of systems. For example, Wynne (2004) details the *ad hoc* checklist used to evaluate online books for the Oxford Text Archive.

Sun and Kantor (2006) explain their experiments with an evaluation approach that distinguishes three levels of granularity and three realities. Their evaluation granularity levels were (1) whether the individual item was retrieved, (2) whether the task-at-hand was completed, and (3) whether the completed task had a valuable impact on the goals-at-hand. These 'granularity' levels represent levels of means to achieve goals, with levels 1-3 ranging from low level means through intermediate goal to higher level goal. Their three realities in evaluation were (1) real users, (2) real systems, and (3) real problems. This model idealizes evaluation approaches that engage real users with real systems and real problems where these approaches attend both the completion of tasks and the achievement of end goals.

2.3 Evaluation in Design Science Research

Little work in the DSR arena has addressed the choice of strategies for evaluation. Walls et al. (1992) introduce the notion of discrete testable hypotheses for explicitly evaluating two components of IS Design theories, the design process and the design product (meta-design) in their ability to achieve meta-requirements. They provide no guidance on how to evaluate, although the presumption seems to be a positivist approach.

March and Smith (1995) emphasize evaluation as one of the two activities in design science: build and evaluate. Evaluation regards the development of criteria and the assessment of the artefact's performance in comparison to the criteria (p. 258). Beyond simply establishing that an artefact worked or didn't work, evaluation also has a responsibility to determine how and why it worked (or not), i.e. using natural science methods for theorizing about IT.

Among their seven guidelines, Hevner *et al* (2004) require researchers to rigourously evaluate design artefacts. They summarize five kinds of evaluation methods (observational, analytical, experimental, testing, and descriptive). However, they do not provide much guidance in choosing among extant evaluation methods.

Evaluation in Computer Science research is similarly developing an emphasis on evaluation, and is also taking a positivist approach, for example in the International Conference on Software Engineering (ICSE) and in the Management of Data (SIGMOD) conferences. For example, Tichy (1998) proposes more use of experiments and Tichy provided a tutorial at ICSE 2007 on

Venable (2006) classified DSR evaluation approaches into two primary forms: artificial and naturalistic evaluation. Artificial evaluation evaluates a solution technology in a contrived and non-realistic way. Naturalistic evaluation explores the performance of a solution technology its real environment i.e., within the organisation. Of these two, Venable notes that naturalistic evaluation is critical, with evaluation in a naturalistic setting being "the real proof of the pudding". Baskerville et al (2007) follow on from this to consider the need for and characteristics of "Soft Design Science Research". The distinction between artificial and naturalistic evaluation relates closely to the three realities described by Sun and Kantor (2006) in that a naturalistic setting would involve *real users* using *real systems* to solve *real problems* (i.e., to accomplish *real tasks* in *real settings*).

Artificial evaluation may be empirical or non-empirical. It is nearly always positivist and reductionist, being used to test design hypotheses (Walls et al 1992). Interpretivist or even critical techniques may also be used, but these generally supplement the main goal of proving or disproving the design theory and/or the utility of the DSR artefact. Artificial evaluation includes laboratory experiments, field experiments, simulations, criteria-based analysis, theoretical arguments, and mathematical proofs. Artificial evaluation is then unreal in some way or ways according to the three realities (Sun and Kantor, 2006), such as unreal users, unreal systems, and especially unreal problems (not held by the users and/or not real tasks, etc.). Importantly, to the extent that an artificial evaluation setting is unreal, evaluation results may not be applicable to real use, thus necessitating naturalistic approaches.

By performing evaluation in a real environment (real people, real systems (artefacts), and real settings (Sun and Kantor, 2006), naturalistic evaluation embraces all of the complexities of human practice in

real organisations. As such, it may be difficult (and costly), partly because it must discern the effects of many confounding variables in the real world. Naturalistic evaluation is always empirical and may be interpretivist, positivist, and/or critical. Naturalistic evaluation methods include case studies, field studies, surveys, ethnography, phenomenology, hermeneutic methods, and action research. To the extent that naturalistic evaluation is affected by confounding variables or misinterpretation, evaluation results may not be precise or even truthful about an artefact's utility or efficacy in real use.

The perspective taken by Venable (2006) is largely ex post, although he accommodates ex ante perspective (e.g. via forms of artificial evaluation, such as simulations or mathematical proofs). Nonetheless, more naturalistic forms of ex ante evaluation are also possible as considered in section 2.2 above.

More recently, Purao and Storey (2008) propose that the Technology Acceptance Model (Davis, 1989) can be used as a predictive theory (Gregor, 2006) to evaluate whether a DSR artefact is likely to be adopted in practice. This non-empirical, artificial evaluation approach is distinctive in that it focuses on the potential efficacy of the IT artefact in that the IT artefact must be adopted in order for there to be an efficacious outcome of IS DSR.

3 FORMULATING A STRATEGIC FRAMEWORK FOR DESIGN SCIENCE RESEARCH EVALUATION

The purpose of this section is to formulate a strategic framework for evaluation in Design Science Research. A strategic framework could serve (at least) two purposes. It could be used to help design science researchers build strategies for evaluation of their research outcomes and to achieve improved rigor in DSR. It could also be used descriptively to improve our understanding of unstated evaluation strategies in extant reports in the DSR literature.

There are several aspects from the above literature that will be valuable in formulating a strategic framework for DSR evaluation. First, the distinction between ex ante and ex post evaluation is useful in identifying the validity of ex ante evaluation, as well as to enhance our understanding of ex post evaluation. It is not absolutely necessary for design researchers to construct an artefact in order to evaluate a design theory. There is a substantial body of work upon which design research may draw in formulating ex ante evaluations.

The literature on evaluation assumes that "ex ante" and "ex post" are unproblematic concepts. However, this literature assumes that a "system" is being evaluated. Ex ante evaluations take place before the system is constructed and ex post evaluations take place after the system is constructed. The system anchors the distinction between ex ante and ex post. Applying these concepts in DSR can become complicated because the system as a reference point anchor may not be relevant. DSR evaluation can instead anchor to the artefact. Thus, ex ante evaluation takes place before the artefact is construction and ex post evaluation takes place after the artefact is constructed.

The artefact as an anchor must itself be clearly defined. Some DSR will regard an IT artefact that is an operating piece of information technology. Design research may regard the design as the artefact, or a process as the artefact. Consequently, it is possible to slide the ex ante – ex post distinction depending upon how the design researchers choose to define their artefact.

Second, the distinction between artificial and naturalistic evaluation is useful. There are advantages to both artificial evaluation (such as more control and lower cost) and naturalistic evaluation (more realism). Evaluation of artefacts in artificial settings is not limited to simple experimental settings, but includes somewhat imaginary or simulated settings where the technology (or its representation) can be studied under substantially artificial conditions. Closely related to this, the design of an evaluation can be manipulated by choosing from among the available realities and measurement granularities (level of goal to be achieved) available to evaluators.

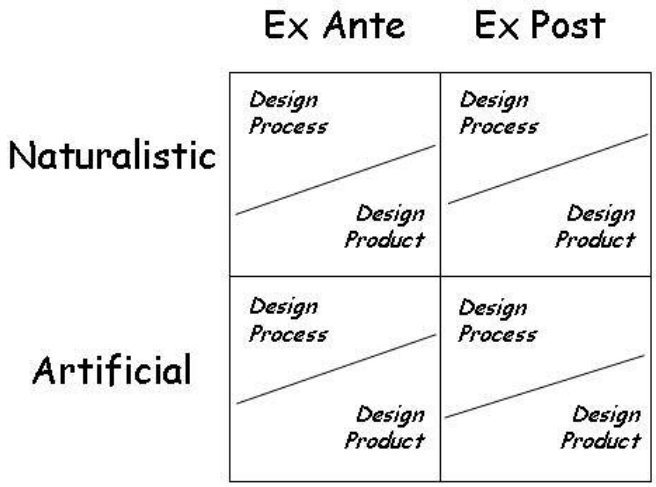

Figure 1. Strategic DSR evaluation framework

Another aspect from the literature that is not surprising is that there is substantial tension between positivism and interpretivism in evaluation. The human determination of value is rather central to this tension, drawing in social, cultural, psychological and ethical considerations that will escape a purely technical-rationality. Note that this relates to granularities and levels of goals identified above, but also to unstated goals and issues of stakeholders.

Drawing upon the above literature, we can formulate a strategic framework by choosing the prominent alternatives described above featuring *when* evaluation takes place, *what* is actually evaluated, and *how* it is evaluated. "When" to evaluate may be selected from ex ante, ex post, or both. It incorporates aspects such as the evaluation context (real users, organizations, and problems). "What" is evaluated may involve choosing between the design process or the design product (*cf.* Walls *et al*, 1992). "How" to evaluate may be selected from naturalistic or artificial forms of evaluation (Venable 2006). See Figure 1.

As noted above, the strategic framework is designed to be used both normatively to advise the design of DSR evaluation and descriptively to understand evaluation in the extant DSR literature. The application of the framework for these two purposes is discussed in the following two sections.

4 APPLYING THE STRATEGIC FRAMEWORK NORMATIVELY

In this section we will describe how to select evaluation strategies using the framework elements. We consider first the ex ante – ex post timing distinction, then issues relating to quality (what to measure) depending on what is evaluated (design product and/or design process).

4.1 Ex Ante versus Ex Post in Design Science Research

Choosing between ex ante or ex post evaluation (or both) in DSR depends on the scope of the research project. The ex ante – ex post concepts prevalent in the evaluation literature fit rather well into the DSR paradigm. Evaluation is not limited to a single activity conducted at the conclusion of a design-construct-evaluate cycle. In fact, there are at least two evaluation episodes available: design-evaluate-construct-evaluate.

If we anchor the evaluation on an IT artefact operating piece of information technology (in the sense of Orlikowski & Iocono 2001), then the evaluation of the design can be regarded as "ex ante" evaluation, before construction of any artefacts. The evaluation of the artefact is usually regarded as "ex post" evaluation, after construction of any artefacts. See Figure 2.

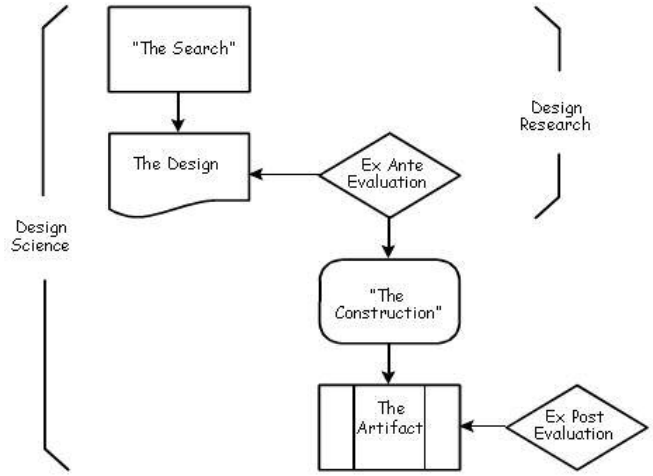

Figure 2. Ex ante versus ex post in DSR

The framing of evaluation as ex ante or ex post helps distinguish between the common meanings of design research and design science. For Simon (1996), the Science of Design is "a body of intellectually tough, analytic, partly formalisable, partly empirical, teachable doctrine" about the design process (p. 113). From this perspective, March and Smith (1995) complete Simon's beginning interest in evaluation, extending the work to construction and evaluation of artefacts. By doing so, they open multiple opportunities for evaluation: particularly prominent are the recognized positions of ex ante (evaluating the design) and ex post (evaluating the artefact).

This distinction assumes that the ultimate goal of DSR is an IT artefact that is an operating technology. For design research, the anchor moves from a constructed piece of technology (the anchor for most DSR) to the design itself (the anchor for design science). In design research being evaluated independent of construction of an operating IT artefact, ex post becomes the evaluation of the design (design being itself an artefact for design research), and ex ante would become the evaluation of the search process. Since our focus in this paper is the broader view of DSR rather than the more narrow scope of design research, we will anchor the ex ante and ex post distinction on the constructed IT artefact in the sense of Orlikowski and Iocono (2001).

4.2 Evaluation Measures: Quality of Design Product and Design Process

Very often the content measures for evaluation of information systems designs and artefacts are closely linked with quality criteria. Quality can be described in terms of more-or-less measurable variables. Differences in quality reflect differences in the quantity or state of some product attribute. The availability of these quality measures clarifies the selection of content measures within the strategic framework. There are many different perspectives to defining the notion of quality (Garvin, 1987). Several authors (cf. Walls et al., 1992) have distinguished between design artefacts that were products (e.g. a new IT system) and processes (e.g. a method for developing an instance of the new type of IT system).

For characterizing DSR evaluation where the design artefact is a **product** we can use a quality model such as ISO 9126 as inspiration. ISO 9126 is an international standard for the evaluation of software. The standard is divided into four parts that address the following subjects: quality model; external metrics; internal metrics; and quality-in-use metrics, together with a structured set of characteristics and sub-characteristics. For example one main characteristic is "Efficiency" which is defined as a "set of attributes that bear on the relationship between the level of performance of the software and the amount of resources used, under stated conditions". The standard suggests a number of potential

measures leading to an evaluation of the design against the characteristics and sub-characteristics of the model. ISO 9126 also offers us quality-in-use measures.

Such measures approximate measures of success (DeLone & McLean, 1992) or user-based definitions. User-based quality definition arises from another school of thought in relation to quality, viz., one in which quality is seen as fitness for intended use.

The user-based definition raises the question: "Who is the user of the design?" For our purposes it also raises the question: "Who is the user of DSR?" If we use Hevner et al.'s (2004) model of DSR we must have at least one user from the business or application domain and one user representing the knowledge base or reference domain

For characterizing DSR evaluation where the design artefact is a **process** we can obtain inspiration from the school of thought around process-based quality. The main idea in process-based quality is that a good process will lead to a good product. A process can be defined as the set of activities, tools, methods, and practices that can be used to guide the flow of production (cf. Humphrey, 1989). Following a well described and sound process yields a better chance of producing quality. Evaluating whether a process is sound is not easy or obvious, but can be done. The components of the process identified above can be evaluated individually against some criteria or opinions of the method/process users can be sought. Finally, the overall efficacy (developing a quality system that satisfies its users, as above) and efficiency (cost and time to develop the system) can also be evaluated.

Hevner et al. (2004) expressed their view on what constitutes good DSR in the form of seven guidelines that are useful for understanding, executing and evaluating design science and design research. For example, their guideline 5 regards rigor in construction and evaluation method. Hevner et al.'s guidelines are to a large extent an example of process-based quality.

5 APPLYING THE STRATEGIC FRAMEWORK DESCRIPTIVELY

The second way in which we can use the strategic framework is as a tool to improve our understanding of the evaluation strategies implicit in reports from the DSR literature. In order to gain insight into DSR research designs, we can apply the framework for interpreting and describing the evaluation strategies implied by published DSR reports. In using the framework descriptively, we ask three main questions, (1) What is actually being evaluated, IT artefact (and if so what exactly) and/or IT development method? (2) How is it being evaluated, naturalistically or artificially, using what process, and evaluating against what criteria? To illustrate this approach, we consider four example DSR publications.

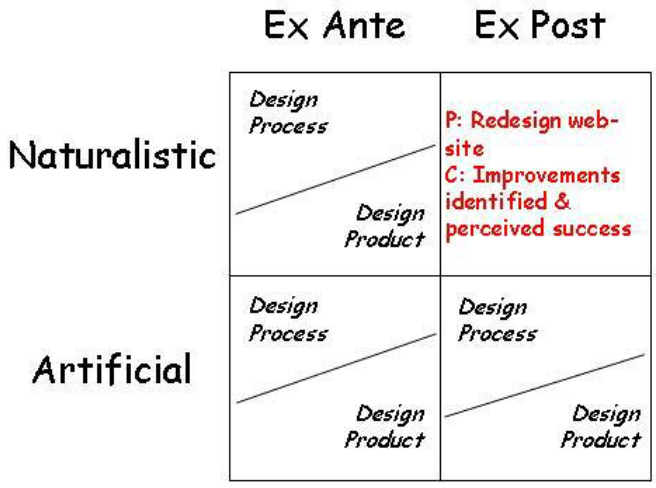

Figure 3. An ex post artificial evaluation strategy (Bell et al., 2007)

In our first example, Bell et al. (2007) develop a design artefact, which is a framework for how one can derive business process semantics from syntactic descriptions of web services. The framework was evaluated using a realistic scenario but "due to the immaturity of the semantic web … the framework could only be experimented on a simulated pilot-project" (p. 74). In Figure 3, we answer the three main questions that our framework proposes (for Bell *et al*, 2007). Figure 3 and subsequent figures, P summarises the essential characteristics of the evaluation *Process*, while C indicates the evaluation *Criteria*.

- *What* is actually evaluated? In Bell et al (2007) it was the framework for deriving semantic business processes from syntactic web services. The result is a design artefact than in itself is a process of how to derive semantics from syntax. Note that the framework
- *How* it is evaluated? The evaluation was an artificial evaluation in which the framework was enacted against a scenario "designed on the basis of previously developed services" (p. 74). As an outcome it "can be assumed that the greater semantic expressiveness of the ontological models would provide benefits …" (p. 82).
- *When* was it evaluated? It was evaluated ex post (after the design artifact was developed).

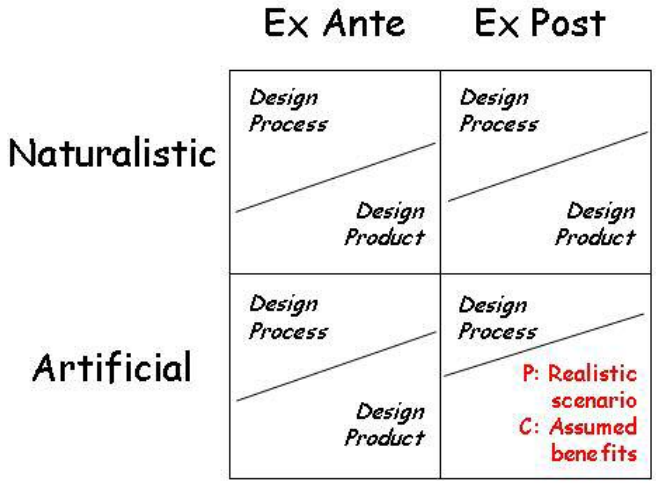

Figure 4. An ex post naturalistic strategy (Albert et al., 2004).

As a second example, using design science, Albert et al. (2004) developed a model called GIST (Gather-Infer-Segment-Track) that can guide the design of web-based systems, as well as the subsequent management of such sites. GIST is in itself a design product that incorporates a process, do 'Gather' before 'Infer' etc. To evaluate the GIST artefact, Albert et al. "observe whether the redesign of the Web site in the business organization resulted in identification of business leads …" (p. 164). The web site being redesigned was in a Fortune 50 company. GIST was applied and the authors "… suggested some design improvements …" (p. 175-176). "This resulted in a tremendous improvement …" (p. 176) and overall "the company considers its new Web site investment and application of GIST a huge success" (p. 178). In Figure 4 we answer the three main questions that our framework proposes (for Albert et al. 2004):

- *What* is actually evaluated? Both the web-site development (a design process) and the web-site management (a design product, albeit a process-oriented one) parts of GIST were evaluated. Thus GIST is both a design product and a design process.
- *How* it is evaluated? A naturalistic evaluation is described in which an existing web-site in a Fortune 50 company was assessed focusing on identifying improvements. Thus it was conducted using a real system in a real organization facing real problems. The study was presumably interpretive with unspecified informants in the company report GIST to be a "huge success".
- *When* was it evaluated? It was evaluated ex post (after the design artifact was developed).

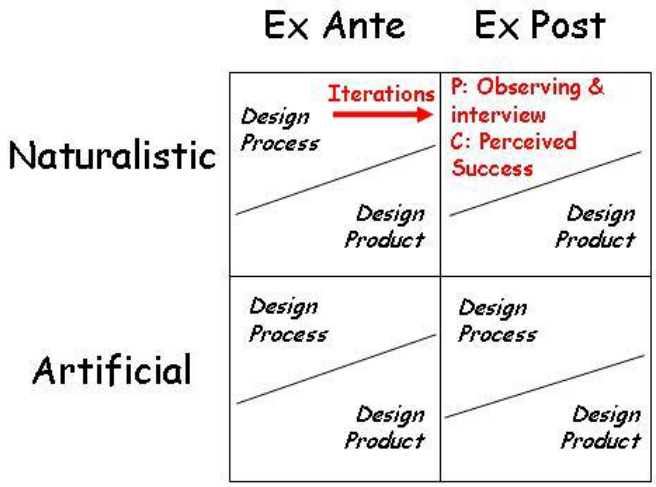

Figure 5. An ex ante - ex post naturalistic evaluation strategy (Arnott, 2006)

In our third example, Arnott (2006) constructed an evolutionary decision support system methodology from a taxonomy of 37 cognitive biases. For evaluation of this design artefact, a naturalistic evaluation method was chosen; Arnott chose to "use the DSS development method in an actual project to evaluate its feasibility and effectiveness" (Arnott, 2006, p. 58). A "single case design" (p. 67) was chosen. The author describes the selection of the case as opportunistic, "in that the actual case was less important than the process being studied". Participant observation was part of the evaluation method as the author was the working as a systems analyst in the DSS development project in the host company (p. 67). Through a number of iterations, a design artefact is developed that takes cognitive biases into account. The evaluation of success was primarily based on the opinion of the managing director; who regarded it as a success (p. 72).

In Figure 5 we answer the three main questions that our framework proposes (for Arnott, 2006):

- *What* is actually evaluated? In this case it was mainly the DDS development process (methodology) that was developed which was being evaluated.
- *How* it is evaluated? A case study which is clearly naturalistic, but with the author as the main developer (participant observer) using the company a case study. Note that the evaluation (research) method could also be considered to be action research.
- *When* was it evaluated? It was evaluated through the iterations described. The main iteration is ex post but some of the earlier iterations may be labelled ex ante.

In our final example, addressing the problem of effective distribution of information, Zhao et al. (2000) examine conventional mailing lists and use the result of that examination to propose a new workflow mechanism. The design consists of a proposal of two new information distribution methods and a proposed extension to existing information filtering algorithms. The paper does not develop a technology artefact as such, but proposes one (or more). The evaluation is therefore clearly ex ante (that is before the building of an artefact). Furthermore there is an artificial evaluation of "a very simple data set based on the seminar announcement in figure 1" (p. 67), where figure 1 in the paper (p. 50) is "an example Seminar Announcement" created specifically for the purpose of the evaluation. In Figure 6 we answer the three main questions that our framework proposes (for Zhao et al, 2000):

- *What* is actually evaluated? A design for a workflow mechanism is evaluated, which can be considered a design product.
- *How* it is evaluated? An artificial evaluation is performed using examples developed solely for the evaluation – but realistic. The contribution is described as the "introduction of a workflow perspective into the domain of information filtering and delivery" (p. 70). The measure used to evaluate is that "it is possible …" (p. 70), i.e. feasibility is demonstrated.
- *When* was it evaluated? It was evaluated ex ante (before any artefact was developed).

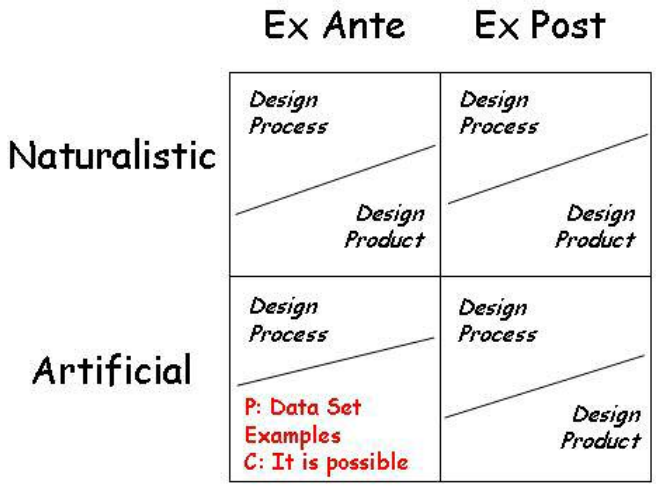

Figure 6. An ex ante artificial evaluation strategy (Zhao et al., 2000).

6 CONCLUSION

Evaluation is a very significant issue in IS DSR. Its importance is widely recognised, yet it is often poorly performed and there is little guidance in the DSR literature concerning how to choose and design an appropriate evaluation strategy.

To address the above need, we have developed and presented an evaluation framework based on an analysis and synthesis of works in IS research and Design Science Research. This framework offers a strategic view of DSR evaluation. The framework includes two dimensions, ex ante vs ex post evaluation and naturalistic vs. artificial evaluation. The ex ante perspective offers the possibility to evaluate prior to undergoing the risk and effort of building an instantiation of the artefact. The ex post perspective offers the possibility of evaluating the instantiated artefact in reality, not just in theory or hypothetically. Naturalistic evaluation methods offer the possibility to evaluate the real artefact in use by real users solving real problems (Sun and Kantor, 2006), while artificial evaluation methods offer the possibility to control potential confounding variables more carefully and prove or disprove design hypotheses, design theories, and the utility of design artefacts.

The primary aim of the new framework is to guide the Design Science researcher. The stage in DSR (early, middle, or late/mature DSR) and the needs/goals of the research (control, applicability to reality) can then be used as inputs to selecting which quadrant in the framework to pursue. Practical issues, (cost, resources, time, etc.) must also be considered in research design, but the framework can then guide understanding of what evaluation is possible within such constraints. More detailed recommendations for evaluation research design choices would need to be based on the literature for each of the particular methods and cannot be addressed in the space allotted for this paper. Thus this remains for future research.

We have shown how the framework can be used descriptively to analyze the evaluation strategy of four published DSR examples. Diverse examples were chosen to illustrate different parts of our framework. This descriptive application demonstrates that there is a wider diversity of evaluations strategies than may be currently assumed by the methodological literature on DSR.

We believe there is an opportunity for using the framework normatively in future research, especially with further development. For example it can be used to surface the evaluation opportunities that present themselves to DSR researchers. Such usage may lead to more comprehensive, rigorous, and/or cost effective strategies for evaluation in future DSR. The framework could also be used by reviewers of DSR publications or research proposals in evaluating research design choices.

References

Albert, T. C., Goes, P. B., & Gupta, A. (2004). Gist: a Model for Design and Management of Content and Interactivity of Customer-Centric Web Sites. *MIS Quarterly, 28*(2), 161-182.

Arnott, D. (2006). Cognitive biases and decision support systems development: a design science approach. *Information Systems Journal, 16*(1), 55-78.

Bannister, F., & Remenyi, D. (2000). Acts of faith: instinct, value and IT investment decisions. *Journal of Information Technology, 15*(3), 231-241.

Bell, D., de Cesare, S., Iacovelli, N., Lycett, M., & Merico, A. (2007). A framework for deriving semantic web services. *Information Systems Frontiers, 9*(1), 69-84.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology, *MIS Quarterly*. 13(3). 319–339.

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research, 3*(1), 60-95.

Garvin, D. (1987). Competing on the Eight Dimensions of Quality. . *Harvard Business Review, 65*(6), 101-109.

Gregor, S. (2006) The Nature of Theory in Information Systems, *MIS Quarterly, vol. 30(3): 611-642.*

Gregor, S. and D. Jones (2007) The Anatomy of a Design Theory, *Journal of the Association for Information Systems*, 8(5) (May), article 2, pp. 312-335.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science In Information Systems Research. *MIS Quarterly, 28*(1), 75-105.

Humphrey, W. S. (1989). *Managing the Software Process*. Boston: Addison-Wesley.

Irani, Z., Love, P. E. D., Elliman, T., Jones, S., & Themistocleous, M. (2005). Evaluating e-government: learning from the experiences of two UK local authorities. *Information Systems Journal, 15*(1), 61-82.

Keast, R., & Waterhouse, J. (2006). Participatory evaluation: a missing component in the sustainable social change equation for public services. *Strategic Change, 15*(1), 23-35.

Klecun, E., & Cornford, T. (2005). A critical approach to evaluation. *European Jn of IS, 14*(3), 229-243.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems, 15*(4), 251-266.

Orlikowski, W. J., & Iacono, C. S. (2001). Research commentary: Desperately seeking "IT" in IT research - A call to theorizing the IT artifact. *Information Systems Research*, 12(2), 121-134.

Purao, S. And Storey, V. C. (2008). Evaluating the adoption potential of design science efforts: The case of APSARA. *Decision Support Systems*. 44. 369–381.

Scarbrough, J. E. (1998). Measuring human life value from the courtroom to the living room. *Journal of the American Society of CLU & ChFC, 52*(1), 68-73.

Shafir, E. B., Osherson, D. N., & Smith, E. E. (1993). The Advantage Model: A comparative theory of evaluation and choice under risk. *Organizational Behavior and Human Decision Processes, 55*(3), 325-378.

Simon, H. A. (1996). *The Science of the Artificial* (3rd ed.). Cambridge, Mass.: MIT Press.

Stockdale, R., & Standing, C. (2006). An interpretive approach to evaluating information systems: A content, context, process framework. *European Journal of Operational Research, 173*(3), 1090-1102.

Sun, Y., & Kantor, P. B. (2006). Cross-Evaluation: A new model for information system evaluation. *Journal of the American Society for Information Science and Technology, 57*(5), 614-628.

Symons, V. J. (1991). A review of IS systems evaluation: Content, context and process. *EJIS, 1*(3), 205-212.

Tichy, W. (1998). Should Compuer Scienteists Experiment More? *IEEE Computer,* 31(5), 32-40.

Vaishnavi, V., & Kuechler, W. (2004). Design Research in Information Systems. Retrieved October 2, 2004, from http://www.isworld.org/Researchdesign/drisISworld.htm

Venable, J. (2006). A Framework for Design Science Research Activities. In *Proceedings of the 2006 Information Resource Management Association Conference* (pp. forthcoming). Washington, DC, USA.

Walls, J. G., Widmeyer, G. R., & El Sawy, O. A. (1992). Building an information system design theory for vigilant EIS. *Information Systems Research, 3*(1), 36-59.

Wynne, M. (2004). Evaluation in the Arts and Humanities Data Service. *VINE, 34*(4), 196-200.

Yang, Y., & Padmanabhan, B. (2005). Evaluation of Online Personalization Systems: A Survey of Evaluation Schemes and A Knowledge-Based Approach. *Journal of Electronic Commerce Research, 6*(2), 112-122.

Zhao, J. L., Kumar, A., & Stohr, E. A. (2000). Workflow-Centric Information Distribution Through E-Mail. *Journal of Management Information Systems, 17*(3), 45-72.