Communications of the Association for Information Systems

Volume 34 Article 10

1-2014

On the Use of Experiments in Design Science Research: A Proposition of an Evaluation Framework

Tobias Mettler
University of St. Gallen, tobias.mettler@unisg.ch

Markus Eurich
Department of Management, Technology, and Economics, ETH Zurich, Switzerland

Robert Winter
Institute of Information Management, University of St. Gallen, Switzerland

Communications of the Association for Information Systems
CAIS

On the Use of Experiments in Design Science Research: A Proposition of an Evaluation Framework

Tobias Mettler
Institute of Information Management, University of St. Gallen, Switzerland
tobias.mettler@unisg.ch

Markus Eurich
Department of Management, Technology, and Economics, ETH Zurich, Switzerland

Robert Winter
Institute of Information Management, University of St. Gallen, Switzerland

Abstract:

Although experiments are considered a valid scientific method for evaluating the outcome of design science research (DSR), only minimal procedural guidelines or standards exist that help researchers in the setup and conduct. To take advantage of and encourage researchers to include design experiments as an inherent part of their build and test cycle, this study proposes a set of guidelines. In order to get a broad overview of how researchers currently apply the experimental method in DSR and to detect potential drawbacks, an extensive review of the extant literature was conducted. On this basis, we propose an evaluation framework that complements the general design science research guidelines of Hevner and colleagues. The purpose of this framework is to assist researchers, reviewers, editors, and readers in understanding possible pitfalls as well as to ask the right questions which need to be answered in the conduct of design experiments.

Keywords: design experiments, design science research, evaluation framework, systematic review

Volume 34, Article 10, pp. 223-240, January 2014

Communications of the Association for Information Systems

I. INTRODUCTION

An increasing amount of research in information systems (IS) is conducted by following the Design Science Research (DSR) approach. Contrary to a behavioral research paradigm, which aims at describing, explaining, and predicting the individual, organizational, or social effects of technology use (or misuse/nonuse), DSR is concerned with "the systematic creation of knowledge about, and with (artificial) design" [Baskerville, Pries-Heje, and Venable, 2008]. With this goal in mind, DSR consists of two major areas: research on design and design as research. The former studies "design" as an intentional, intellectual, and creative activity for problem solving [Chandrasekaran, 1990], while the latter sees it as an entity or artifact which surpasses the boundaries of human and organizational capabilities [Hevner, March, Park, and Ram, 2004]. Typically, a scientifically rigorous design process comprises multiple iterations of (at least) two main activities: (1) building, and (2) evaluating the solution to the previously identified problem. According to Hevner et al. [2004], this last step is "crucial" for demonstrating the utility, quality, and efficacy of a designed artifact. Thus, the evaluation step serves as a "catalyzer" for assessing whether a new artifact veritably expands the general knowledge base of the field. Evaluation comprises both: testing the solution against the defined requirements and assessing its actual impact in reality.

Many authors argue for a pluralistic use of artificial and naturalistic evaluation methods such as action research, case studies, simulation, prototyping, or informed arguments [Cole, Purao, Rossi, and Sein, 2005; Pries-Heje and Baskerville, 2008]. Since DSR aims at producing generic knowledge, it is assumed that design experiments are an ideal technique for testing newly developed artifacts and for systematically deriving design improvements. In the seminal work of Hevner et al. [2004] the importance of experimental research is also highlighted by it being included in the five proposed design evaluation methods.

Although experiments are being considered as a valid research method for evaluating the quality of the design process and the design product under development [Wohlin, Runeson, Höst, Ohlson, Regnell, and Wesslén, 2000], they, up until now, have been applied only rarely in DSR. For instance, in the sub-area of decision support systems, it was found that only 4.1 percent of the analyzed papers utilized controlled experiments for evaluations. Dominant approaches include simulation, scenarios, and case studies, which were found with a relative frequency of 20.4, 15.7, and 11.6 percent, respectively in the analyzed papers [Arnott and Pervan, 2008]. A similar distribution can be assumed in other areas of IS research.

A possible reason for disregarding experimental evaluation in DSR may be the absence of well-defined and widely accepted guidelines for conducting, interpreting, and presenting design experiments and their results [Kitchenham, Pflegger, Pickard, Jones, Hoaglin, El-Emam, and Rosenberg, 2002; Perry, Porter, and Votta, 2000]. Moreover, in contrast to other research disciplines, "experimental design" is often not included in the methodology courses in the curriculum of an IS student [Adelman, 1991]. A particularly problematic fact is that the existing experimental protocols are based primarily on assumptions of behavioral research and are often not sufficient or specific enough for design-oriented research.

Therefore, the major objective of this article is to provide an overview of how experiments with design are currently being conducted and, based on this analysis, to deduce an evaluation framework for IS researchers and educators who are planning on teaching such a methodology course.

II. DESIGN SCIENCE RESEARCH AS APPROACH TO PRACTICAL PROBLEM SOLVING

According to Simon [1996, p. xii], the sciences of the artificial "are concerned not with the necessary but with the contingent—not with how things are but with how they might be—in short, with design." As a science, design science has "to do with the systematic creation of knowledge about, and with, design" [Baskerville, 2008, p. 441]. From this central idea, different connotations, denotations, and definitions of what constitutes the central scientific outputs of DSR were discussed in recent literature [Carlsson, 2006; Hevner and Chatterjee, 2010; Iivari, 2007; McKay and Marshall, 2007; Offermann, Blom, Schönherr, and Bub, 2010; Orlikowski and Iacono, 2001; Winter, 2008].

Design Research Artifacts

The development and evaluation of specific problem solutions, at times also referred to as *artifact research* [Horvath, 2001], has a special emphasis on sufficiently new or decisively better "artifacts," constituting the "possibly sole, or chief, output of the research" [Gregor and Jones, 2007, p. 318]. The term *artifact* is used to describe something that is artificial, or constructed by humans, in contrast to something that occurs naturally [Simon, 1996]. A widely accepted breakdown of outputs is to distinguish among four artifact types [March and Smith, 1995]: (1) constructs, (2) models, (3) methods, and (4) instantiations. In the recent literature, design theories are also seen as another possible result of design research (e.g., Baskerville and Pries-Heje, 2010; Goldkuhl, 2004; Gregor and Jones, 2007; Kuechler and Vaishnavi, 2008b). The above artifact types are defined in the following paragraphs.

Following Hevner et al. [2004, p. 78], "*constructs* provide the language in which problems and solutions are defined and communicated ...." They form the abstract foundation for describing the problem space as well as for conceptualizing the solution space. In some instances they may be used to represent the entire shared knowledge of a discipline or sub-discipline [March and Smith, 1995]. However, the "construct" artifact can have varying levels of formalism. Examples of highly formalized constructs include algebraic data specification languages [Chuang and Chien, 2005], lexical semantics [Basili and Pennacchiotti, 2010], and modeling primitives [Liu, Bhattacharya, and Wu., 2007]. Furthermore, informal constructs, like natural language, can be used to describe the problem and solution space. However, this may increase the ambiguity of the conceptualization and even generate confusion [Theodoulidis and Loucopoulos, 1991].

Models generally represent a formal description of "... some aspects of the physical or social reality for the purpose of understanding and communicating" [Mylopoulos, 1992, p. 51]. Depending on the notion of the representation, they are either descriptive (i.e., they give an unprejudiced reproduction of some aspects of reality), explanatory (i.e., they depict causal links in order to better understand reality), or prescriptive (i.e., they recommend an efficient solution state of a future reality). In DSR, models are usually descriptive; they consider different constructs of a particular application domain in relation to one another. Models may be applied for documenting both the problem (initial state) and the solution (desired state). Exemplary "model" artifacts are process models [Ralyté and Rolland, 2001], enterprise architecture models [Winter and Fischer, 2007], or maturity models [Mettler and Rohner, 2009].

In contrast to models, "method" artifacts are used to solve a problem "... based on a specific way of thinking, consisting of directions and rules, structured in a systematic way in development activities with corresponding development products" [Brinkkemper, 1996, p. 275]. Consequently, methods are systematic (i.e., they give rules on how to act and instructions on how to solve problems), goal-oriented (i.e., they define standards on how to proceed or act in order to reach a defined goal), and repeatable (i.e., they are inter-subjectively practicable). Contrary to descriptive models, in which the problem and solution state descriptions are rather static, methods are explanatory and dynamic in such a way as to provide a direct plan to systematically change from a problem to a solution state. In doing so, methods often rely on formal models. Examples of "method" artifacts are software development methods [Rossi, Ramesh, Lyytinen, and Tolvanen, 2004] and management design methods [Pries-Heje and Baskerville, 2010].

The "instantiation" artifact is the most complex research output, as it is the operationalization or physical conversion of prior solution components. According to March and Smith [1995, p. 258] it can also be seen as "… the realization of an artifact in its environment." In doing so, they may be used as a proof-of-concept to demonstrate the feasibility and effectiveness of the models and methods they contain [March and Smith, 1995]. The current scientific-knowledge-base comprises a wide number of instantiations, including, for example, systems for expert applications [Kuechler and Vaishnavi, 2008a], decision support [Pries-Heje and Baskerville, 2008], and business intelligence [Chung, 2006].

Artifacts are not exempt from natural laws or behavioral theories and often rely on "existing kernel theories that are applied, tested, modified, and extended through the experience, creativity, intuition, and problem solving capabilities of the researcher" [Hevner et al., 2004, p. 76]. Hence, the second strand of DSR is focused on the principles of design and general systematization of theoretical design knowledge (e.g., Gregor, 2009; Markus, Majchrzak, and Gasser, 2002; Pries-Heje and Baskerville, 2010; Purao, 2002; Walls, Widmeyer, and El Sawy, 1992). The motivation for considering design theories as a fifth major research output is twofold: first, developing theories is required in order to make a sounder basis for vindicating the rigor and legitimacy of DSR (e.g., Gregor and Jones, 2007; Venable, 2006); second, there is an inherent need for "theories for design and action" or "narrative thinking" to complement traditional prescriptive theories in order to make artifact construction more efficient (e.g., Goldkuhl, 2004; Gregor, 2006; Kuechler and Vaishnavi, 2008b). Iivari [2007, p. 41] argues that the main difference between behavioral theories and artifacts (to which we also assign design theories) is the applicability of the concept of truth: "While it is meaningful to speak about truth or "truthlikeness" in the case of (behavioral) theories, it is not in the case of artifacts. Artifacts are only more or less useful for human purposes." Examples of such design theories are Kasper's [1996] theory of decision support system design, Walls et al.'s [1992] theory for vigilant executive information systems, Markus et al.'s [2002] for emergent knowledge processes, and Brohman, Piccoli, Martin, Zulkernine, Parasuraman, and Watson's [2009] theory for strategic network-based customer service systems.

Design Research Process

In accordance with Archer [1969] the logical nature of the act of designing does not necessarily have to be dependent on the artifact which is to be designed. Several researchers, therefore, have tried to systematize a common DSR process by decomposing and classifying the central design activities and by finding rationales for effective problem solving (e.g., Coyne, 1988; Nunamaker, Chen, and Purdin, 1990; Takeda, Veerkamp, Tomiyama, and Yoshikawa, 1990). However, a commonly agreed upon procedure model still does not exist. The extant literature suggests breaking down a common DSR process into four mutually dependent yet methodologically distinct steps: (1) problem analysis, (2) artifact construction, (3) artifact evaluation, and (4) interpretation, theory construction, and learning.

According to Purao [2002], the design process is usually initiated by a “need and requires intention” as part of the problem analysis. However, at the beginning of the search process, the requirements and applicability of a solution are usually quite fuzzy, which leads to a state of obscurity and a potential misfit between the needs and the goal [Wieringa, 2009]. A sound analysis of the problem or “innovation opportunities” [Järvinen, 2007], thus, is considered an essential first step in the design process in order to define the relevance and orientation (e.g., Peffers, Tuunanen, Gengler, Rossi, Hui, Virtanen, and Bragge, 2006; Takeda et al., 1990; Vaishnavi and Kuechler, 2008).

By definition, the core of a DSR process is to design a purposeful artifact to address a (previously identified) relevant problem [Nunamaker et al., 1990]. Artifact construction has to be carried out in a transparent and traceable way, demonstrating practical feasibility on the one hand, and possessing methodological validity on the other [Hevner et al., 2004]. Following “the logic of design” [Simon, 1996], a researcher has to stick with the solution development until a satisfactory alternative is found.

Whether a solution is satisfactory or not is determined during the evaluation step. The phases of a DSR process are interdependent and constitute a diaphanous cycle of building and testing an artifact according to established criteria of relevance and rigor [March and Smith, 1995]. Hevner et al. [2004] suggest a plurality of different methods for design evaluation, such as case studies, controlled experiments, simulation, prototyping, and informed arguments. Nevertheless, the broad margin of applicable methods not only brings more freedom, but also more difficulties in assessing the results with respect to criteria of relevance and rigor. Cole et al. [2005] particularly emphasize the use of action research, since it shares the common philosophy of pragmatism. In addition, we perceive design experiments as an instrumental means for evaluating artifacts, due to the positivistic consciousness of many design science researchers.

III. EXPERIMENTS AS A MEANS OF EVALUATING DESIGN ARTIFACTS

In this study the analysis is restricted to experiments which emphasize the evaluation of design alternatives or so-called “candidate solutions” [Su, Hakkarainen, and Brasethvik, 2004] with respect to their impact on reality. Therefore, we introduce the notion of “design experiments” as a synonym for a scientific method that facilitates causal inference while testing the effects of different design alternatives by adhering to the principles of (1) *control,* (2) *randomization*, and (3) *manipulation*. A design experiment is understood to be an evaluation procedure for gaining generalized knowledge about the utility of distinct *design alternatives* by following the aforementioned principles. The overall objective is to ensure the inter-subjective verifiability of a superior design for a particular application domain and to find a way to improve the design product and process.

Control refers to the presence of a control group. In an experimental setting there are two groups: a test group and a control group. The differences between the test group and the control group are the basis for differentiating between reality and wishful thinking and for testing for cause-and-effect relationships. Only the test group is exposed to a stimulus (i.e., the new or better design), but not the control group. After the experiment the reactions of the participants of the test group are compared to those of the control group.

Randomization means that the subjects (e.g., participants, data sets, etc.) are randomly assigned to either the test and or the control group; i.e., they are not already involved in the project or emotionally attached to the solution design. In this way the prior existent differences between the participants are reduced and the testing results are more stable, because known and unknown confounding factors are distributed in a uniform way among the test users.

Manipulation in a design experiment means that the artifact (e.g., a construct, model, method, or instantiation) is evaluated under differing conditions (cf. Parikh, Javid, Sasikumar, and Ghosh, 2006). The terms *independent* and *dependent* variable are often used to distinguish between the parts of the artifact, which are actively changed (independent variable) and the state changes, which are measured under each condition (dependent variable) [Boring, 1933]. In a controlled environment only the independent variables are subject to deliberate manipulations.

The investigator assesses the effect a change the independent variable has on the dependent variable. Experimentation differs from pure observation because it requires the researcher to actively "manipulate" the cause(s) of a phenomenon and control the test environment.

These three principles need to be carefully addressed in the three basic phases of a design experiment: pre-experimental phase, actual experimental phase, and analysis phase. Prior to the experiment ("planning" in Figure 1), the researcher must create the experimental procedure, which includes the following steps: First, the randomization technique as well as the required sample size is defined. Then the researcher determines the treatments and manipulation procedure that the participants will undergo and the metrics with which the dependent variable will be measured. Finally the researcher predefines how the empirical data is to be analyzed. In an experimental design process, there should be at least one other researcher who actually conducts the experiment and who is unbiased and trained to perform the tests ("conducting" in Figure 1). This is done in order to avoid the so-called Pygmalion effect or Rosenthal effect [Rosenthal and Jacobson, 1968] during the experiment and to avoid any kind of bias in the data analysis. In the pre-experimental phase this researcher's task is to randomly assign the participants to either the test group or the control group. In the actual experiment the researcher conducts the experiment, i.e., he gives the instructions and exposes the participants to different treatments. Finally, the researcher calculates and interprets the results after the experiment. Figure 1 is revisited in the guidelines section.

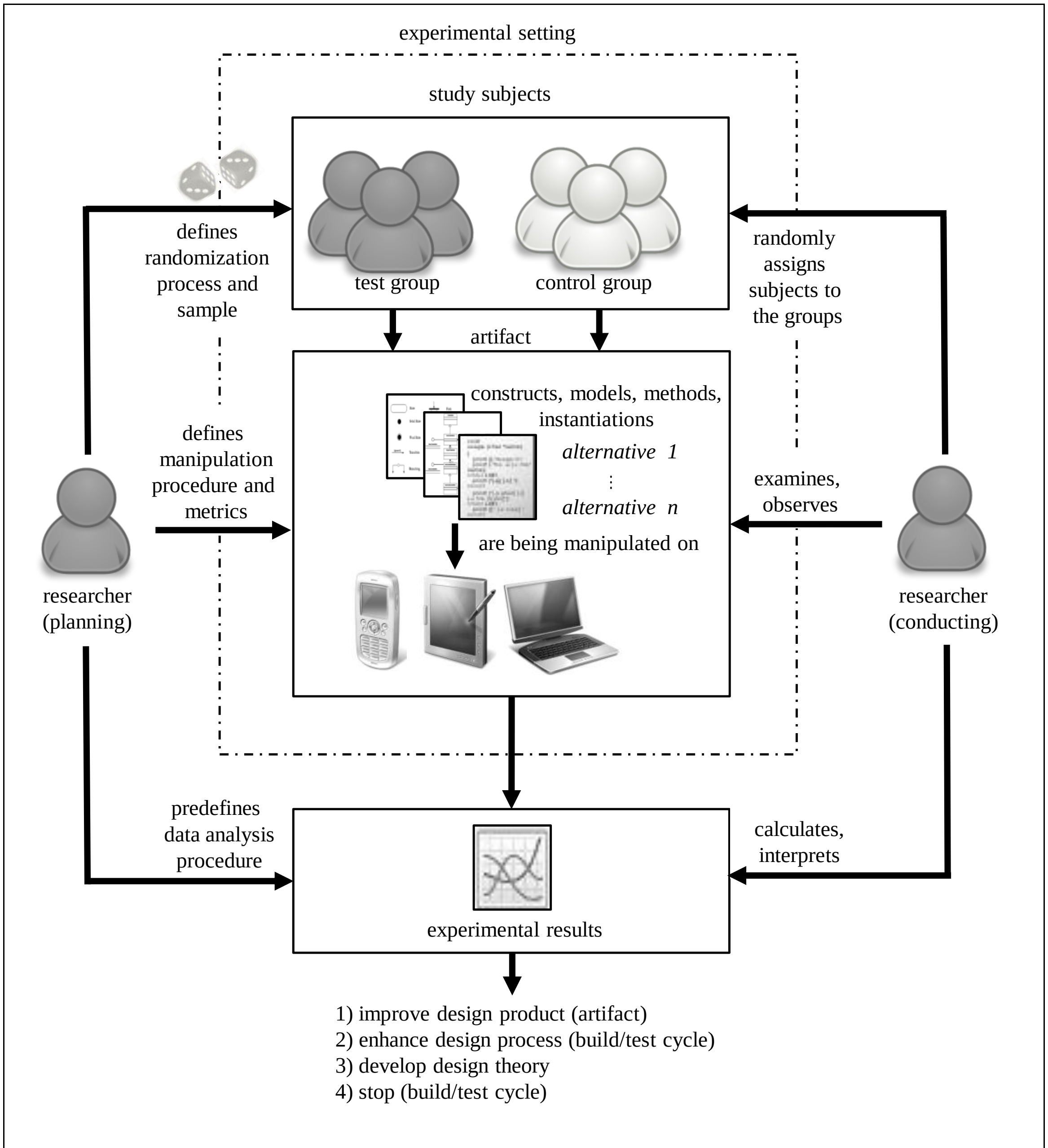

Figure 1. Mental Model of Experimentation with Design

Differences Between Natural or Social Sciences Experiments and Design Experiments

Design experiments are one way to systematically evaluate the effects of different designs or the overall qualities of an artifact. The main aim of doing such experiments is to identify the most useful problem-solving alternative. Depending on the complexity of the artifact, the design experiment may include controlling and monitoring a large number of factors. This is fundamentally different from natural or social sciences experiments, which are rather focused on understanding causal relationships under controlled conditions. The ultimate goal here is to find truthful links among a few defined and controllable variables.

Another difference can be found with respect to the manipulation or treatment of variables. While natural or social sciences experiments typically modify one variable at a time, a sequenced alteration is not always possible in the case of design experiments. Depending on the complexity of the artifact, a lot of parallel interactions are possible, which is not easy for the researcher to control. In order to increase control, the main strategy of design researchers is to simplify the designs or reduce manipulation alternatives, whereas natural and social sciences researchers focus more on the selection and assignment of participants in order to reduce extraneous effects and to increase control. As a result, reproducing design experiments is extraordinarily difficult and possible only if an exact study protocol and the artifact are available. Reproducing natural and social sciences experiments is feasible if the exact experimental protocol is available.

Finally, another major difference can be found with regard to well-defined and widely accepted guidelines for conducting, interpreting, and presenting the results of experiments. While a broad range of technical quality criteria exists for natural or social sciences experiments, to our knowledge no guidelines are available for design experiments. Hence, with the conceptualization of the experiment, a design science researcher always needs to develop quality criteria that are helpful to other researchers for understanding the validity and reliability of the design experiment.

Table 1: Differences Between Natural/Social Sciences Experiments and Design experiments

CHARACTERISTICS	NATURAL OR SOCIAL SCIENCES	DESIGN SCIENCE
Aim	Controlled testing to understand causal relationships (truth)	Controlled testing to evaluate qualities of an artifact (utility)
Number of studied factors	Typically few variables	Typically a large number of studied factors
Manipulation	Typically one interaction at a time	Typically many parallel interactions at a time
Control of environment	Reduction of extraneous variables by selecting and assigning study participants	Reduction of extraneous variables by simplifying the design or reducing the manipulation alternatives
Reproducibility	Feasible when study protocols are available	Feasible when study protocols and artifacts are available
Reporting results	Broad range of technical quality criteria available	No standard quality criteria available (must be developed with experiment)

Current Use of Design Experiments in IS

In order to investigate how design science researchers conduct experiments, we chose a systematic review approach. This approach is usually understood to be an explicit, rigorous, and transparent methodology for analyzing information exhibited in research publications [Greenhalgh, Robert, McFarlane, Bate, and Kyriakidou, 2004]. As the documentation of research outcomes often varies significantly in both content and format, it is a crucial step within a systematic review to correctly conduct the search process. In order to do so, there are two fundamental strategies [Webster and Watson, 2002]: (1) A thematic analysis (i.e., searching for papers according to previously defined themes or topic maps), and (2) a bibliometric analysis (i.e., including/excluding publications based on measureable quantitative and/or qualitative criteria).

In order to assess the nature and the strengths and weaknesses of experimentation in DSR, we applied a mixed approach. We restricted our search to research articles presented at major conferences between 2005 and 2011, which is the period of time after the dissemination of the seminal DSR guidelines by Hevner et al. [2004]. We also limited our analysis to publications that truly used design experiments (and not misleading other research designs such as case study or observation) for the purpose of evaluating newly developed IS/IT artifacts (and not for an ex-post analysis of existing artifacts). This was necessary since a previous search for finding articles in top-ranked journals which explicitly applied design experiments for evaluating their artifacts was unsuccessful.

In order to identify a first set of publications specifically related to experimentation in DSR, the following two databases were queried, using the search strings “design science research,” “design science,” “experiment,” and “experimentation” (and its various combinations) in the titles and keywords: ACM Digital Library (ACM), with the intention of finding papers that use experimental approaches to evaluate first and foremost the technical quality of designed artifacts (e.g., performance, security, usability), and AIS Electronic Library (AIS) to find papers that used experimental approaches for assessing rather organizational use aspects of designed artifacts (e.g., cost-benefit ratio, cultural fit, task-technology fit). Our queries yielded ninety-seven publications, where eighty-six papers were found in ACM and eleven papers in AIS. After applying the aforementioned exclusion criteria (i.e., no reference to evaluation or a misconception of experimental research design), a total of fifty publications were analyzed in more detail. The analysis included the reading of each paper’s abstract, methodology, and results sections, as well as eliciting the metrics discussed in the next section.

The coding of our representative sample of research publications was organized to analyze the number and percentages of the sources, the articles arising each year, the scientific publication outlets in which they appeared, the number of authors involved, their experimental setting, the type of experiment, the particular technique applied for data gathering, the qualities of subjects used as data sources, and the objective of the experiment. This systematic manner of data capturing was extremely labor-intensive, but “it has the advantage that it can illuminate the deep structure of the field in a way that is impossible to achieve with other literature analysis approaches” [Arnott and Pervan, 2008].

We used open, axial, and selective coding techniques, as discussed by Urquhart [2001], to categorize more “fuzzy” information such as the actual objective of a design experiment or the deficiencies encountered in the experimental design. The coding was done independently by two coders, which allowed for posterior cross checking and elimination of discrepancies or unclear codes. Both information forms (i.e., quantitative numbers and open coded categories) serve as a basis for the proposition of an evaluation framework.

Sample

According to the results of our coding procedure, design experiments are most frequently conducted in the field of human–computer-interaction. The majority (20 percent) of the articles were published at the ACM Conference on Human Factors in Computing Systems (CHI). We also identified two articles (accounting for 4 percent each) at the ACM International Conference on Multimedia (MM) and the ACM Symposium on Applied Computing (SAC). The majority of the analyzed articles (72 percent) were occasionally published at other conferences, e.g., at the IFIP WG 8.2 and 8.6 International Working Conference, the Participatory Design Conference, and the International Conference on Wireless, Mobile, and Ubiquitous Technology in Education. Hence, no particular outlet was detected that privileged the publication of design experiments. However, we found that particularly in IS-related communication bodies, such as the European Conference on Information Systems and the International Conference on Information Systems, experiments for evaluating artifacts are rarely present.

Characteristics of Design Experiments in IS

In order to assess how design experiments are conducted, we analyzed the type of experiment, experimental setting, the technique for data gathering, and the subjects who build the data basis for such studies (Figure 2).

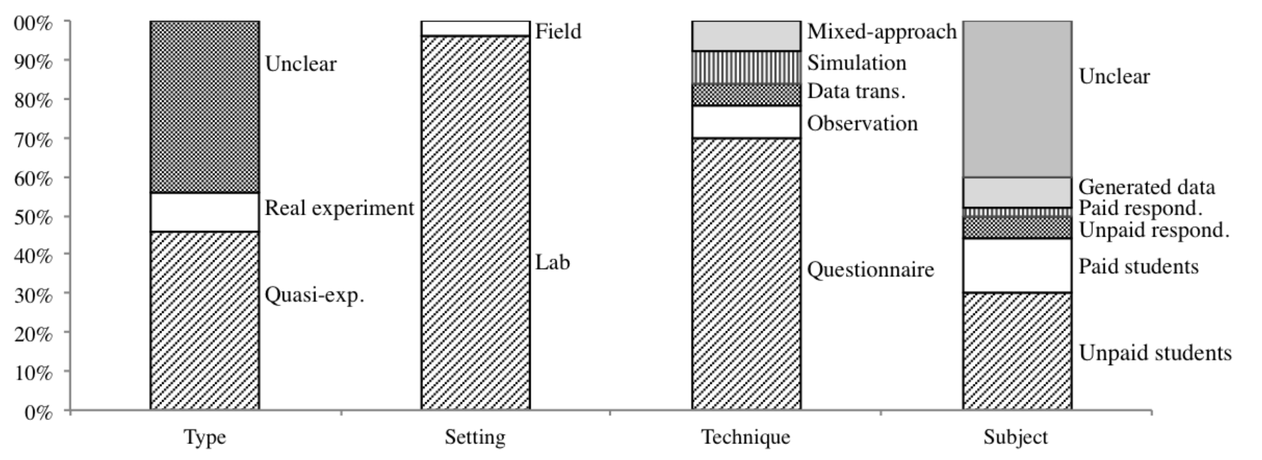

Figure 2. Characteristics of Design Experiments in IS

We found that the vast majority of design experiments were conducted in a laboratory environment (96 percent). Only two studies (4 percent) evaluated their developed artifacts in the field. Around 46 percent of the identified papers were codified as quasi-experiments since the studies lacked controlled scheduling of data collection procedures (e.g., use of randomization, control groups). Only about 10 percent of the papers fulfilled the requirements of "true" experiments. In 44 percent of the papers it was not possible to clearly categorize the type of experimental design since no information about the sampling and data-gathering technique is revealed.

The majority of design experiments (70 percent) use a questionnaire for data acquisition. In doing so, basically two modes are in use: "demonstration" of the innovative artifact (e.g., new visualization technique) and the "interaction" of subjects with the innovative artifact (e.g., use of a new retrieval algorithm). The experiences or impressions of the subjects are then used to answer the questions on a questionnaire. Instead of letting the users answer a questionnaire for obtaining data about the qualities of the developed artifact, around 8 percent of the studies rely on observation (e.g., protocolling a reaction of a subject to a particular stimulus that was generated by the artifact). Only a few studies (6 percent) use data transmission instruments such as sensors or eye tracking for recording "physically" observable data. An additional 8 percent of the papers renounce the use of human judgment to assess the qualities of the artifacts and applied simulation techniques as an alternative. Another 8 percent used a mixed approach, adopting several of the aforementioned techniques for data acquisition. With respect to the sample or data basis used for evaluating newly developed artifacts, we found twenty-two articles (44 percent) that rely on the appraisal made by students. In fifteen of the twenty-two papers, the students participated on a voluntary basis, and, in the remaining seven papers, the students' efforts were refunded. In 6 percent of the identified studies, unpaid, independent respondents served as data basis. Only one paper (2 percent) bases the artifact evaluation on answers from compensated respondents. The papers that use simulation as the preferred means for evaluating the artifact (8 percent) do not rely on subjects and use generated data. In twenty articles (40 percent) the data source for assessing the artifact remains unclear.

Evaluating the quality of a newly developed artifact can be performed in various ways. As it is often not possible to assess all of the potential benefits and drawbacks of a new solution (e.g., due to limited time, funding, personnel, or restricted access to respondents) [Baskerville et al., 2008], a design experiment needs to be clearly focused on one particular aspect which is to be evaluated (e.g., in form of defined hypotheses). With our coding procedure, we identified six main categories for distinguishing the objectives of design experiments in IS (Figure 3).

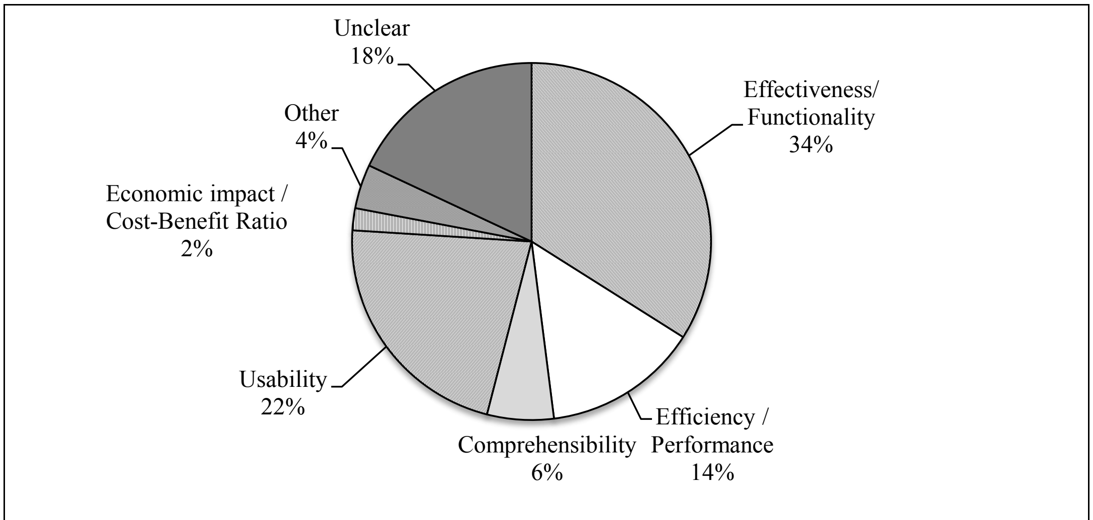

Figure 3. Objectives of Design Experiments in IS

We found that 34 percent of the identified papers focus on the evaluation of an artifact's effectiveness (e.g., testing the fit with the initially defined functional and non-functional requirements). Fourteen percent of the articles emphasize the assessment of the artifact's efficiency (e.g., assessing its performance in comparison with another solution). In 6 percent of the papers, the main objective of the design experiment is to measure the increase/ decrease of understandability (e.g., comparing the new modeling notation with existing ones), and in another 22 percent the assessment of usability is the central concern (e.g., testing the accessibility of content for people with impairments). Only one paper (2 percent) focuses on assessing the economic impact of the newly developed artifact. An additional 4 percent of the articles aim at other objectives. In 18 percent of the cases, it remains unclear what the authors intended to test, because the measures are operationalized only rudimentarily and defined hypotheses or research questions are missing.

To evaluate the usability and utility of an artifact, a great part of the analyzed studies relied on users who interact with or practically apply the developed solution [Brohman et al., 2009]. When humans are involved in the design experiment, a certain level of subjectivity often is introduced, which is commonly referred to as informant bias [Winter, 2010]. A single informant bias may emerge if the statements about the qualities of an artifact are collected from one single source (e.g., "key informant") or a homogenous group (e.g., "power users"), a group that shares a common educational background, personal position, or knowledge deficiency. Since many of the analyzed papers did not sufficiently characterize its used sample, it's hard to say if an informant bias is a general issue in design experiments. However, it is clearly a shortcoming to not clearly specify the "sources" of the obtained data.

In addition, it is also not possible to determine whether the design experiments are subject to a population bias (i.e., a distortion of results due to a difference between the experimentally accessible population and the target population). Different sampling and randomization techniques exist to reduce a potential bias. Since many of the studies did not include randomized assignments but rather relied on "available" IS/IT-students to build up their experimental sample, an indication for a systematic deficiency is given (e.g., students may have a higher affinity to technology than the actual real users, but they may lack practical experience). At least, we found that randomization processes are infrequently described in the analyzed papers. Similarly, the selection process of the sample is often not discussed.

Most experiments were conducted in a laboratory environment. We assume that this was done in order to improve control and to reduce confounding factors. Nevertheless, we cannot clearly say if the analyzed design experiments really applied different control variables or control groups (e.g., pretest/posttest control groups) in order to reduce errors. An ecological bias can also be frequently seen in design experiments. These are errors which stem from "incorrectly describing independent variable(s), incorrectly describing or measuring dependent variable(s), multiple-treatment interference, interaction of history and treatment effects, interaction of time of measurement and treatment, pretest and posttest sensitization" and so forth [Chin, 2001]. Again, only a few studies addressed that issue.

Finally, another weak point is that researchers often fail to clearly construct their hypotheses. While constructing hypotheses has been commonplace in other fields such as the natural sciences or social sciences since, for a long time, the analyzed studies showed a clear deficit in constructing process and product hypotheses in a formally correct manner. Furthermore, we found that dependent variables, which are not directly observable (such as the usability or effectiveness of a designed artifact), are often not operationalized in a reproducible manner. In a couple of papers it was not clear what the goal of the design experiment was, nor was it clear what was measured or how the measurement was done in order for the authors to arrive at their conclusions. Moreover, complex constructs, such as usability, are measured by just one operationalized variable (e.g., like/dislike of the artifact), and, therefore, do not capture the full spectrum of the objective which is to be measured.

IV. PROPOSITION OF AN EVALUATION FRAMEWORK FOR DESIGN EXPERIMENTS

The literature analysis shows that there are shortcomings in both planning and conducting experimental design studies. For instance, we found that many of the analyzed publications did not clarify the overall aim, the concrete setting, or the procedure and metrics that were applied in the design experiment. This imprecision makes it difficult for fellow researchers to fully comprehend the contribution of and the further knowledge development based on the design experiment results.

Following a pragmatic epistemology [Goles and Hirschheim, 2000; Goldkuhl, 2012; Wicks and Freeman, 1998], we propose an evaluation framework that supports both researchers in conducting design experiments and reviewers of DSR experimental studies in posing the right questions (cf. Figure 4).

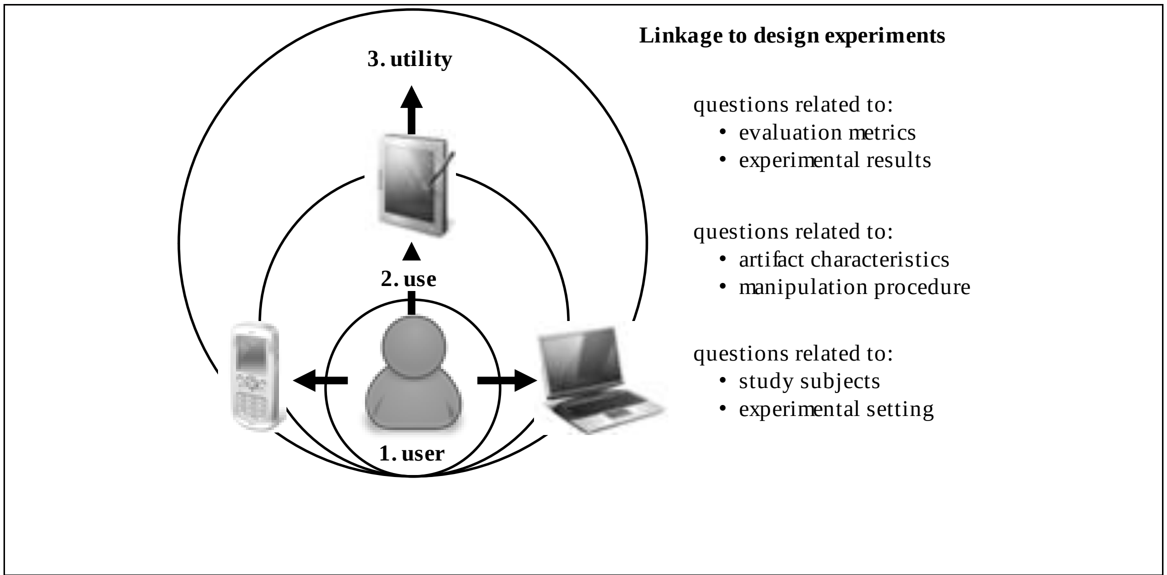

Figure 4. Proposed Evaluation Framework

User

The centerpiece in our evaluation framework is the *user*. The user determines how an artifact is used (or, in the view of the creators, "misused") and what value she can gain from it. By seeing DSR as utility-driven and by using a pragmatic research approach, a design experiment is deemed successful when users judge an artifact as being superior with respect to technological, aesthetic, economical, ethical, or other aspects. It is worth noting that these utility beliefs can strongly differ depending on the users' capabilities, the context in which the user is active, and the situation in which a user applies the artifact. As discussed before, it is hard to speak about truthful results, but, instead, the results from a design experiment need to be put into perspective, more precisely, into the perspective of the user (or user group) for which the artifact is developed. In this sense, the user should be the bottom line for the evaluation.

The user is also an important concept in behavioral IS research, where the determinants of IS acceptance and continued IS use are examined, such as in Davis's (1989) technology acceptance model (TAM) and successive models. These are based on the users' reasoned appraisal of pre-adoption beliefs with regard to the perceived usefulness, perceived ease of use, and the assumption that usage behavior (or system usage) is a consequence of this appraisal. However, as opposed to the mentioned behavioral theories where a high-level conceptualization of the "user" concept often is applied, it is of importance in DSR so as to design experiments, to have a clear understanding of the users' capabilities, habits, working environments, etc., in order to build and test a useful artifact.

> *Example: In one study a design experiment was conducted to investigate the impact of two distinct business process modeling notations on the learning outcome of novice business analysts or software developers.*

This example shows what we found in some of the analyzed articles, i.e., students were used as proxies for professional business users. Although this is particular problematic in studies that wish to replicate behavior of top executives or senior professionals, it might not be devastating for every experiment (e.g., evaluation of consumer-oriented artifacts). A researcher conducting a design experiment, hence, need to ask himself/herself whether students or other substitutes of real users genuinely are suitable for reproducing user behavior or for acting as key informants in utility assessments with respect to a developed artifact.

As with any experimental study, it is also crucial to describe how the users were selected for the design experiment. The researchers have to decide whether they want to apply a randomized process to assign specific users to either the test or the control group (referred to as "real experiment," high internal validity) or to assign users manually, e.g., by matching strategies (e.g., by considering age, business role, and/or computer literacy; referred to as a "quasi-experiment").

Besides a description of user characteristics and sampling procedure, it is also important to circumscribe the experimental setting, especially when relying on user feedback (instead of simulated data) for conducting the design experiment. In the previous example, the sample of the experiment consisted of postgraduate students with knowledge in one notation (and with absolutely no knowledge of the other). This example illustrates two issues: First, it might be debatable how reasonably the results can be transferred from postgraduate students' perception to business analysts, who are the targeted user group for the process modeling tool. Second, there might be a possible conflict of interest because the students might sense pressure or a certain preference of their instructor toward a certain modeling notation. This might influence how students perceive the problem-solving capacity of the distinct business process modeling notations. In this sense, a clarification of the experimental setting (environment, selection process, possible conflicts, etc.) might have helped to reduce a possible disbelief of the explanatory power of the design experiment. Especially the declaration of *conflict of interests* has largely been discussed in other fields. For instance, it has become a compulsory requirement in clinical studies for many years and should also be considered by design researchers, as it could enhance the trustworthiness of the design experiment's results (e.g., Medical Research Council, 2012; National Health and Medical Research Council of Australia, 2007).

Use

Apart from the user, it is essential to provide detailed information about the *use* of the artifact: it is not only the characteristics of the user that can influence the artifact's utility, but the situation and how the artifact is used are also important. For example, the improvement of a scheduling algorithm might be a good solution in the case of unsolved, complex decision-making scenarios; however, it is less useful in circular, rather simplistic decision-making scenarios where already proven heuristics exist. While conducting design experiments and communicating the results, it is important to shed some light on the environment in which the artifact will be used (what the artifact is good for, what not) as well as on the intended scope and goals of the evaluation itself. Defining and explicitly describing the goals and the scope of usage helps the researchers stay focused, because a design experiment has many parallel interactions and influencing factors (cf. Table 1). In general, there are two distinct types of *goals* of a design experiment [Järvinen, 2007]:

> *Goal type I*: A proof that a very new artifact works, i.e., a solution to an unsolved problem was found. It has to be specified why the problem is actually a problem and why no solutions have existed so far. Examples include testing whether the defined problem is solved, as well as how well it is solved. Design experiments may help to identify gaps in the problem-solving abilities of the new solution and may provide indications for further improvements.
>
> *Goal type II*: A proof that the designed artifact works better than existing solutions, i.e., the artifact solves a problem much more efficiently and with fewer resources. It needs to be determined why existing solutions do not fully solve the problem, or why an improved solution is needed. Examples include feature comparison, performance tests, usability tests, and economic analyses. Reasons why existing solutions were not able to fully solve the problem can include former bandwidth restrictions, the prior impossibility of miniaturization, only recent cost decrease of data transfer, and the introduction of new standards.

In the case of goal type I experiments, a design researcher is confronted with a serious dilemma: The newer or more innovative a solution is, the greater is the difficulty in observing the functioning of the artifact in a real environment with real users [Mettler, 2011]. Furthermore, alternative solutions with which to compare the new artifact might be inexistent. Accordingly, a field experiment relying on user feedback as evaluation basis might then not be a good option (cf. Section V; situations in which experimentation with design is not the best approach). In such a use scenario, for instance, experiments with simulated data would be a better choice.

In goal type II experiments, it is of particular importance to provide a description of the meta-requirements [Walls et al., 1992], i.e., "a class of goals or contexts to which a potential problem solution applies." As discussed in the previous section, the problem-solving ability and utility of an artifact is a relative concept (e.g., designed for a dedicated group of users, companies, or domain; emphasizing a dedicated design goal such as cost, quality, flexibility, or efficiency; compared against other existing solutions). The specified meta-requirements, therefore, help to define the scope to which the results of the design experiment can provide meaningful insights. In defining the meta-requirements, moreover, it is not sufficient to simply state that the artifact should be more efficient than existing solutions, but it should be specified in which terms the artifact is supposed to be more efficient, e.g., in response time, ease of use, costs (cf. next subsection). Many of the identified articles lacked a clear description of use scenarios or "meta-requirements."

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

In this example the objective of the experiment is to prove that the newly designed artifact works better than existing solutions (goal type II), namely that the new way of visualizing data works better than comparable existent visualizing methods. The value of the visualization of data is a relative concept; i.e., it depends on MBA students' prevailing habits, computer literacy, job position, performance expectancy, effort expectancy, and trust. A much more precise scheme to define the scope of design experiments could have looked like: "developing a new method for visualizing data of type α for user type β in domain γ with the aim of improving goal δ and comparing it with a traditional method against problem solving capacity of goal δ."

Another fundamental challenge of design experiments is to describe how the artifact actually is manipulated by users. Reproducibility and re-runs of design experiments (e.g., in other environments, at other points in time, with other users, etc.) are crucial for the validity of results. While it is a good starting point to describe the overall setting of the design experiment (cf. previous subsection), it is still not enough to ensure that a design experiment is reproducible. In order that other researchers are able to reconstruct the design experiment, a precise description of the procedure of how the artifact is being used (e.g., in form of test scripts, scenarios, flowcharts) and how data is collected and analyzed (e.g., step-by-step description, validation protocols) is required. It is fundamental to clearly specify how researchers and users interact with the artifact under evaluation. In addition, the procedure of capturing possible reactions, e.g., to changing artifact states (in case of a user-based experiment) or data points (in case of a simulation-based experiment), must be specified.

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

Although the artifact characteristics and metrics for evaluating the new method are extensively discussed in this study, it is still difficult for other researchers to replicate this design experiment with another solution, since no specific procedural protocol is provided. Such a protocol could have described how participants were briefed to use the additional information in an existing programming environment. Test scripts would have helped to answer the questions, "What programming problems were asked?" "How were these questioned asked?" "In which situations were the students allowed to use the new method?" and "What other auxiliary materials did the students have?" A concrete description of who (e.g., moderating researcher, respondents) does what (e.g., perform predefined test case), when (e.g., triggered by a specific event), where (e.g., replicating a stressful situation), and how (e.g., find deficiencies without compiling) is required in order to ensure that the design experiment is reproducible by other researchers. With this call for extensively describing the manipulation procedure of artifacts, there is also the need for providing researchers with the necessary means to do so. For instance, creating new opportunities for publishing research protocols, such as is common in medicine (e.g., JMIR Publications, 2013; Nature Protocols, 2013), or providing a platform for disseminating and interchanging artifacts. Particularly the last point is critical, since a full replication of the design experiment is possible only when the artifact is made available to other researchers.

Utility

Finally, the last element in the evaluation framework is *utility*. Utility emerges through the use of the artifact and is dependent on the user and the environment. In order to operationalize the assessment of an artifact's utility, clear and measurable variables (also referred to as metrics) are needed. The great majority of the analyzed articles measured the utility in rather technical terms, such as the degree to which their artifact is superior in response time, throughput, availability, or latency in comparison to another artifact. In addition, there is a magnitude of other possible aspects that may provide some evidence that the newly developed artifact is useful (cf. Tullis and Albert, 2010). Regardless of the prevailing notion to comprehend "utility," it is important to clearly define the applied evaluation criteria or metrics in order to allow other researchers to replicate the experiment. This also helps to foster the understanding of potential trade-offs that were made during the design of the artifact (e.g., ease-of-use vs. performance). Since DSR addresses "wicked problems" (Hevner et al., 2004) (i.e., unstable requirements and environments, complex interactions, dependence upon human cognitive, and social abilities), additional metrics have to be formulated that can measure mediating and moderating effects on the results of the design experiment. Typical examples of such influences are user-specific characteristics (e.g., age, gender, and computer literacy), date and time effects (e.g., differences of bandwidth utilization on specific workdays), technical effects (e.g., divergent behavior of the designed artifact on different platforms), environmental effects (e.g., divergent behavior of the designed artifact because of temperature differences), and socio-cultural effects (e.g., assignment of distinct connotations and meanings for the same artifact construct because of a different cultural background). In the end, the definition of relevant metrics for quantifying the utility of a particular type of artifact may also represent a valuable contribution itself.

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

Mediating and moderating effects play a pivotal role in this case example. For instance, the "goodness" of hits may be estimated differently by the users depending on contextual and situational factors. Hence, general statements, such as "algorithm A is better than algorithm B," are hard to validate since the utility of an artifact is a relative concept. Moreover, when comparing two artifacts (e.g., a new solution against an existing solution), it is important not only to describe metrics that are observable at both ends, but also to find measures that give equal consideration to the underlying problem the distinct artifacts want to address. For instance, when an artifact A is designed for functionality (or business users) and artifact B is designed for joyfulness (or hedonic users), metrics have to account for both design rationales (cf. Park and El Sawy, 2008); otherwise, the results of the study may lead to one-sided, possibly equivocal conclusions.

V. SITUATIONS IN WHICH EXPERIMENTATION WITH DESIGN IS NOT THE BEST APPROACH

According to Hevner et al. [2004], evidence from design science research must address two major questions: "What utility does the new artifact provide?" and "What demonstrates that utility?" Whereas the first question is in reference to the relevancy of the problem, the second question hints rather to the adequacy and rigor of the solution. From our point of view, a design experiment is a commendable approach for answering the second type of question. However, since design science deals with "wicked problems," an experimental evaluation may not always lead to the inter-subjective results one is hoping for: the higher the level of abstraction and the more complex the artifact is, the more difficult it is to design an experiment where it is possible to manipulate all possible scenarios. Especially for *socio-technical* artifacts, other evaluation methods, such as case-study, ethnography, or action research, may provide much richer insights into the practicality and utility of the designed artifact.

Another important aspect that deserves to be mentioned is that reproducing design experiments can be difficult. In general there are three central prerequisites for such an experiment: (i) The environmental conditions and the research setting must be absolutely identical for the test and the control group, (ii) the people in both the test and the control group must be taken from exactly the same sample, and (iii) the only difference between the test and control is the treatment to which the test group is exposed. Unlike in natural sciences where the "building blocks" of an experiment (e.g., atoms, simple molecules, complex organic compounds) are tangible by using appropriate instruments, the bits and pieces of a human-made artifact can be intangible and of inverse nature. Even if the procedures for manipulation, data gathering, and data analysis are well described, a complete reproducibility of the exact steps is extremely improbable (unless researchers start interchanging their designed artifacts with others). Again, the greater the complexity of the artifact under study is, the more entry points for confounding influences exist.

Moreover, following Järvinen [2007], design science research can yield either totally new results, similar results to earlier ones, or the results that contradict the results of earlier experiments. In the first case (artifact for unsolved problems), it is generally difficult to use design experiments, since only one alternative artifact may exist but several alternatives would be needed in order to really conduct the experiment (cf. Figure 1). The difficulty in the remaining two cases (artifact for solved problems) is in accessing the existing alternatives. Although a researcher may know that a problem solution already exists in the DSR knowledge base, she/he may not be able to gain access to that artifact and, thus, will have to replicate or imitate an existing artifact. This introduces an additional bias into the study which is practically impossible to remove.

From a philosophical and epistemological perspective it must also be mentioned that an analytical-nomological position underlies the experiments. However, this position does not perfectly qualify for all kind of research endeavors: for instance, the dialectical-critical approach postulates that scientific research cannot study the facts detached from the social reality. Whenever an emancipatory and critical assessment of the facts is important for the quality of the outcome, a design experiment needs to be complemented with a critical social and self-refectory assessment.

Finally, design experiments also require advanced capabilities from the investigators. The transition process from a building activity to an evaluating activity is particularly difficult and may demand some changes in the mindset of the researcher in terms of philosophical assumptions (i.e., adaptation from solution search to theory/solution testing). It is thus only recommendable, if the researchers have a rich and long-term experience with design science problem solving.

VI. CONCLUDING REMARKS

The major advantage of experiments is its ability to explore state changes of artifacts in a systematical way, which is hardly achievable with pure observation or measurement. The high level of systematization makes the experimental approach one of the best methods to appraise the utility of a newly developed artifact. As a consequence, the findings can be transferred to the DSR knowledge base more effectively and serve as a profound basis for practical decisions. Yet the experimental approach loses much of its credibility if the experiments are not conducted with the necessary methodological rigor.

The literature review revealed that this is partially the case in some of the analyzed articles. A comprehensive understanding of the necessity of carefully designing and consequently applying methodological rigor in experiments has not yet found its way into the mindset of all DSR researchers: characteristics of users testing the design alternatives are rarely described, the experimental setting is unclear, manipulation procedure and metrics are frequently not operationalized in a reproducible manner. These methodological deficiencies lead to a poor credibility and reliability of the findings, which devalues the perceived quality and significance, as well as the applicability of the experimental approach. To unfold the big advantages and strengths of the experimental approach, the experiments must be conducted in rigorous methodological way.

We consider the proposed evaluation framework as supplementary assistance for design researchers willing to conduct experiments as part of their artifact evaluation strategy or for educators willing to teach design experiments in IS methodology courses. The proposed framework enriches the current DSR literature by reflecting on the current state of the use of design experiments in IS and by expatiating the relevant aspects and questions a researcher has to answer so that the evaluation of the designed artifact is reproducible and credible for other scientists and practitioners. We think that especially novice researchers and students may benefit from the proposed framework. The article also makes a contribution to the clarification of the term *experiment* and its relation to and relevance for design research. Although experiments in design research have been frequently alluded to [Hevner et al., 2004; Kuechler and Vaishnavi, 2011], a definition was still missing.

As with any empirical study, this work has limitations. The primary limitation is the restricted set of keywords we used as a starting point for our search process. Certainly a broader set of terms, including, for example, specific artifacts types, would have yielded more results on which to build our analysis. A broader search would have probably added some journal publications to our list of analyzed articles, as the current keywords did not return any journal articles. The second limitation is the level of detail of our framework. When planning and conducting design experiments, there are undoubtedly many complexities and dilemmas that must be foreseen. Our framework can provide only rough guidance in addressing these problems. Although we tried to propose some pragmatic measures to deal with certain criticisms and problems, we cannot render detailed tips for a particular design experiment. Researchers still need to translate the framework ideas to the respective context of a particular design experiment.

Further research and discussions are needed to formulate artifact or domain specific methodological guidelines for conducting, interpreting, and presenting the results of design experiments. This would be the basis for researchers and educators of DSR to conduct and report on practically relevant, methodologically and theoretically strong experimental results that are accepted throughout and across the IS field.

REFERENCES

Editor's Note: The following reference list contains hyperlinks to World Wide Web pages. Readers who have the ability to access the Web directly from their word processor or are reading the article on the Web, can gain direct access to these linked references. Readers are warned, however, that:

1. These links existed as of the date of publication but are not guaranteed to be working thereafter.
2. The contents of Web pages may change over time. Where version information is provided in the References, different versions may not contain the information or the conclusions referenced.
3. The author(s) of the Web pages, not AIS, is (are) responsible for the accuracy of their content.
4. The author(s) of this article, not AIS, is (are) responsible for the accuracy of the URL and version information.

Adelman, L. (1991) “Experiments, Quasi-experiments, and Case Studies: A Review of Empirical Methods for Evaluating Decision Support Systems”, *IEEE Transactions on Systems, Man, and Cybernetics,* (21)2, pp. 293–301.

Archer, L.B. (1969) “The Structure of the Design Process”, in Broadbent, G., and A. Ward (eds.), Design Method in *Architecture*, London: Lund Humphries, pp. 76–102.

Arnott, D., and G. Pervan (2008) “An Assessment of DSS Design Science Using the Hevner, March, Park and Ram Guidelines”, in Gregor, S., and D.N. Hart (eds.), *Information Systems Foundations: The Role of Design Science,* Canberra, Australia: ANU E Press, pp. 255–284.

Basili, R., and M. Pennacchiotti (2010) “Distributional Lexical Demantics: Toward Uniform Representation Paradigms for Advanced Acquisition and Processing Tasks”, *Natural Language Engineering,* (16)4, pp. 347–358.

Baskerville, R. (2008) “What Design Science Is Not”, *European Journal of Information Systems,* (17)5, pp. 441–443.

Baskerville, R., J. Pries-Heje, and J. Venable (2008) “Evaluation Risks in Design Science Research: A Framework”, V. Vaishanvi and R. Baskerville (eds.), *Proceedings of the 3rd International Conference on Design Science Research in Information Systems and Technology,* Atlanta, GA, pp. 329–334.

Baskerville, R.L., and J. Pries-Heje (2010) “Explanatory Design Theory”, *Business & Information Systems Engineering,* (2)5, pp. 271–282.

Boring, E.C. (1933) *The Physical Dimensions of Consciousness*, New York: The Century Company.

Brinkkemper, S. (1996) “Method Engineering: Engineering of Information Systems Development Methods and Tools”, *Information and Software Technology,* (38)4, pp. 275–280.

Brohman, M., G. Piccoli, P. Martin, F. Zulkernine, A. Parasuraman, and R. Watson (2009) “Design Theory Approach to Building Strategic Network-based Customer Service Systems”, *Decision Sciences,* (40)3, pp. 403–430.

Carlsson, S.A. (2006) “Design Science Research in Information Systems: A Critical Realist Perspective”, *Proceedings of the 17th Australasian Conference on Information Systems,* Adelaide, Australia.

Chandrasekaran, B. (1990) “Design Problem Solving: A Task Analysis”, *AI Magazine,* (11)4, pp. 59–71.

Chin, D.N. (2001) “Empirical Evaluation of User Models and User-adapted Systems”, *User Modeling and User-Adapted Interaction,* (11)1–2 pp. 181–194.

Chuang, S.-L., and L.-F. Chien (2005) “Taxonomy Generation for Text Segments: A Practical Web-based Approach”, *ACM Transactions on Information Systems,* (23)4, pp. 363–396.

Chung, W. (2006) “Designing Web-based Business Intelligence Systems: A Framework and Case Studies”, *Proceedings of the 1st International Conference on Design Science Research in Information Systems and Technology,* Claremont, CA, pp. 147–171.

Cole, R., S. Purao, M. Rossi, and M.K. Sein (2005) “Being Proactive: Where Action Research Meets Design Research” *Proceedings 26th International Conference on Information Systems,* Las Vegas, NV, pp. 325–336.

Coyne, R. (1988) *Logic Models of Design,* London: Pitman.

Goldkuhl, G. (2004) “Design Theories in Information Systems: A Need for Multi-grounding”, *Journal of Information Technology Theory and Application,* (6)2, pp. 59–72.

Goldkuhl, G. (2012) “Pragmatism vs Interpretivism in Qualitative Information Systems Research”, *European Journal of Information Systems,* 21(2), pp. 135–146.

Goles, T., and R. Hirschheim (2000) “The Paradigm Is Dead, The Paradigm Is Dead, Long Live the Paradigm: The Legacy of Burell and Morgan”, *Omega*, 28(3), pp. 249–268.

Greenhalgh, T., G. Robert, F. McFarlane, P. Bate, and O. Kyriakidou (2004) “Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations”, *The Milbank Quarterly,* (82)4, pp. 581–629.

Gregor, S. (2006) “The Nature of Theory in Information Systems”, *MIS Quarterly,* (30)3, pp. 611–642.

Gregor, S. (2009) “Building Theory in the Sciences of the Artificial”, *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology,* Philadelphia, PA.

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

Hevner, A.R., and S. Chatterjee (2010) *Design Research in Information Systems: Theory and Practice,* New York: Springer.

Hevner, A.R., S.T. March, J. Park, and S. Ram (2004) "Design Science in Information System Research", *MIS Quarterly,* (28) 1, pp. 75–101.

Horvath, I. (2001) "A Contemporary Survey of Scientific Research into Engineering Design", *Proceedings of the 13th International Conference on Engineering Design,* Glasgow, United Kingdom, pp. 13–20.

Iivari, J. (2007) "A Paradigmatic Analysis of Information Systems as a Design Science", *Scandinavian Journal of Information Systems,* (19)2, pp. 39–63.

Järvinen, P. (2007) "On Reviewing of Results in Design Research", *Proceedings of the 15th European Conference on Information Systems,* St. Gallen, Switzerland, pp. 1388–1397.

JMIR Research Protocols (2013) http://www.researchprotocols.org (current October 2013).

Kasper, G.M. (1996) "A Theory of Decision Support System Design for User Calibration", *Information Systems Research,* (7)2, pp. 215–232.

Kitchenham, B., S. Pflegger, L. Pickard, P. Jones, D. Hoaglin, K. El-Emam, and J. Rosenberg (2002) "Preliminary Guidelines for Empirical Research in Software Engineering", *IEEE Transactions of Software Engineering,* (28)8, pp. 721–734.

Kuechler, B., and V. Vaishnavi (2008a) "An Expert System for Dynamic Re-coordination of Distributed Workflows", *Expert Systems with Applications,* (34)1, pp. 551–563.

Kuechler, B., and V. Vaishnavi (2008b) "On Theory Development in Design Science Research: Anatomy of a Research Project", *European Journal of Information Systems, (*17)5, pp. 489–504.

Kuechler, B., and V. Vaishnavi (2011) "Promoting Relevance in IS Research: An Informing System for Design Science Research", *Informing Science: The International Journal of an Emerging Transdiscipline,* (14), pp. 125–138.

Liu, R., K. Bhattacharya, and F. Wu (2007) "Modeling Business Contexture and Behavior Using Business Artifacts", *Proceedings of the 19th International Conference on Advanced Information Systems Engineering,* Trondheim, Norway, pp. 324–339.

March, S.T., and G.G. Smith (1995) "Design and Natural Science Research on Information Technology", *Decision Support Systems,* (15)4, pp. 251–266.

Markus, M.L., A. Majchrzak, and L. Gasser (2002) "A Design Theory for Systems That Support Emergent Knowledge Processes", *MIS Quarterly,* (26)3, pp. 179–212.

McKay, J., and P. Marshall (2007) "Science, Design, and Design Science: Seeking Clarity to Move Design Science Research Forward in Information Systems", *Proceedings of the 18th Australasian Conference on Information Systems,* Toowomba, Australia, pp. 604–614.

Medical Research Council (2012) "Good Research Practice: Principles and Guidelines", http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC002415 (current August 2012).

Mettler, T. (2011) "Maturity Assessment Models: A Design Science Research Approach", *International Journal of Society Systems Science,* (3)1/2, pp. 81–98.

Mettler, T., and P. Rohner (2009) "Situational Maturity Models as Instrumental Artifacts for Organizational Design", *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology,* Philadelphia, PA.

Mylopoulos, J. (1992) "Conceptual Modelling and Telos", in Loucopoulos, P., and R. Zicari (eds.), *Conceptual Modelling, Databases and Case,* New York: John Wiley & Sons, pp. 49–68.

National Health and Medical Research Council of Australia (2007) "Australian Code for the Responsible Conduct of Research", http://www.nhmrc.gov.au/guidelines/publications/r39 (current July 2007).

Nature Protocols (2013) http://www.nature.com/nprot/index.html (current October 2013).

Nunamaker, J.F., M. Chen, and T.D.M. Purdin (1990) "Systems Development in Information Systems Research", *Journal of Management Information Systems,* (7)3, pp. 89–106.

Offermann, P., S. Blom, M. Schönherr, and U. Bub (2010) "Artifact Types in Information Systems Design Science: A Literature Review", *Proceedings of the 5th International Conference on Design Science Research in Information Systems and Technology,* St. Gallen, Switzerland, pp. 77–92.

Orlikowski, W.J., and C.S. Iacono (2001) “Research Commentary: Desperately Seeking ‘IT’’ in IT research—A Call to Theorizing the IT Artifact”, *Information Systems Research,* (12)2, pp. 121–134.

Parikh, T., P. Javid, K. Sasikumar, K. and P. Ghosh (2006) “Mobile Phones and Paper Documents: Evaluating a New Approach for Capturing Microfinance Data in Rural India”, *Proceedings of the 2006 ACM Conference on Human Factors in Computing Systems,* Montreal, Canada, pp. 551–560.

Park, Y., and O.A. El Sawy (2008) “Towards a Design Theory for Hedonic Systems: Delivering Superior User Experience in the Digital Home Entertainment Context”, *Proceedings of the 3rd International Conference on Design Science Research in Information Systems and Technology,* Altanta, GA.

Peffers, K., T. Tuunanen, C.E. Gengler, M. Rossi, W. Hui, V. Virtanen, and J. Bragge (2006) “The Design Science Research Process: A Model for Producing and Presenting Information Systems Research”, *Proceedings of the 1st International Conference on Design Science in Information Systems and Technology,* Claremont, CA, pp. 83–106.

Perry, D.E., A.A. Porter, and L.G. Votta (2000) “Empirical Studies of Software Engineering: A Roadmap”, *Proceedings of the Conference on The Future of Software Engineering,* Limerick, Ireland, pp. 345–355.

Popper, K. (2002) *The Logic of Scientific Discovery, second English edition,* New York: Routledge.

Pries-Heje, J., and R. Baskerville (2008) “The Design Theory Nexus”, *MIS Quarterly,* (32)4, pp. 731–756.

Pries-Heje, J., and R. Baskerville (2010) “Management Design Theories”, *Proceedings of the International Working Conference on Human Benefit through the Diffusion of Information Systems Design Science Research,* Perth, Australia, pp. 263–281.

Pries-Heje, J., R. Baskerville, and J. Venable (2008) “Strategies for Design Science Research Evaluation”, *Proceedings of the 16th European Conference on Information Systems,* Galway, Ireland, pp. 255–266.

Purao, S. (2002) “Design Research in the Technology of Information Systems: Truth or Dare”, GSU Department of CIS Working Paper, Atlanta, GA.

Ralyté, J., and C. Rolland (2001) “An Assembly Process Model for Method Engineering”, *Proceedings of the 13th International Conference on Advanced Information Systems Engineering,* Interlaken, Switzerland, pp. 267–283.

Rosenthal, R., and L. Jacobson (1968) “Pygmalion in the Classroom”, New York: Holt, Rinehart & Winston.

Rossi, M., B. Ramesh, K. Lyytinen, and J.-P. Tolvanen (2004) “Managing Evolutionary Method Engineering by Method Rationale”, *Journal of the Association for Information Systems,* (5)9, pp. 356–391.

Simon, H.A. (1996) *The Sciences of the Artificial, third edition,* Cambridge, MA: MIT Press.

Su, X., S. Hakkarainen, and T. Brasethvik (2004) “Semantic Enrichment for Improving Systems Interoperability”, *Proceedings of the 2004 ACM Symposium on Applied Computing,* Nicosia, Cyprus, pp. 1634–1641.

Takeda, H., P. Veerkamp, T. Tomiyama, and H. Yoshikawa (1990) “Modeling Design Processes”, *AI Magazine,* (11)4, pp. 37–48.

Theodoulidis, C.I., and P. Loucopoulos (1991) “The Time Dimension in Conceptual Modelling”, *Information Systems Journal,* (16)3, pp. 273–300.

Tullis, T., and B. Albert (2010) *Measuring the User Experience,* Burlington: Morgan Kaufmann.

Urquhart, C. (2001) “An Encounter with Grounded Theory: Tackling the Practical and Philosophical Issues”, in Trauth, E. (ed.), *Qualitative Research in IS: Issues and Trends*, London: Idea Group Publishing, pp. 104–140.

Vaishnavi, V., and B. Kuechler (2008) *Design Science Research Methods and Patterns: Innovating Information and Communication Technology,* Boca Raton, FL: Auerbach Publications.

Venable, J. (2006) “The Role of Theory and Theorising in Design Science Research”, Chatterjee, S., and A.R. Hevner (eds.), *Proceedings of the 1st International Conference on Design Science Research in Information Systems and Technology,* Claremont, CA, pp. 1–18.

Walls, J.G., G.R. Widmeyer, and O.A. El Sawy (1992) “Building an Information System Design Theory for Vigilant EIS”, *Information Systems Research,* (3)1, pp. 36–59.

Webster, J., and R.T. Watson (2002) “Analyzing the Past to Prepare for the Future: Writing a Literature Review”, *MIS Quarterly,* (26)2, pp. xiii–xxiii.

Wicks, A.C., and R.E. Freeman (1998) “Organization Studies and the New Pragmatism: Positivism, Anti-positivism, and the Search for Ethics”, *Organization Science,* 9(2), pp. 123–140.

Wieringa, R. (2009) “Design Science as Nested Problem Solving”, *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology,* Philadelphia, PA.

Winter, R. (2008) “Design Science Research in Europe”, *European Journal of Information Systems,* (17)5, pp. 470–475.

Winter, R. (2010) “Organisational Design and Engineering: Proposal of a Conceptual Framework and Comparison of Business Engineering with Other Approaches”, *International Journal of Organisational Design and Engineering,* (1)1/2, pp. 126–147.

Winter, R., and R. Fischer (2007) “Essential Layers, Artifacts, and Dependencies of Enterprise Architecture”, *Journal of Enterprise Architecture,* (3)2, pp. 7–18.

Wohlin, C., P. Runeson, M. Höst, M. Ohlson, B. Regnell, and A. Wesslén (2000) *Experimentation in Software Engineering: An Introduction,* Dordrecht, The Netherlands: Kluwer Academic Publisher.

ABOUT THE AUTHORS

Tobias Mettler is an Assistant Professor at the Institute of Information Management, University of St. Gallen (HSG), Switzerland, where he is leading the Competence Center Health Network Engineering. His research interests are in the area of design science research, technology adoption, and organizational learning, with a particular focus on health systems and services. Dr. Mettler is the principal and co-investigator of several large national and international funded research projects in the area of technology, medicine, and management. His work has been published in *Communications of the Association for Information Systems, Decision Support Systems, Health Services Management Research,* and *Health Systems.* Before joining HSG, he worked several years in industry research and in management consulting.

Markus Eurich is a Postdoctoral Research Fellow in the Department of Management, Technology, and Economics at ETH Zurich. He completed his Diplom (MSc equivalent) in Management Information Systems from the University of Mannheim and his Ph.D. from ETH Zurich and conducted research at the University of Cambridge. His research focuses on the design and transformation of business models, especially in the sector of information and communication technology. He led and participated in several projects on designing business models and service innovation, high performance computing, and technology and innovation management. Prior to his academic career, he worked several years in the IT industry, mainly in the fields of Business Consulting and Customer Relationship Management.

Robert Winter is tenured chair of Business & Information Systems Engineering at University of St. Gallen (HSG), director of HSG’s Institute of Information Management, founding academic director of HSG’s Executive Master of Business Engineering program and academic director of HSG’s Ph.D. in Management program. He received master degrees in business administration and business education as well as a doctorate in social sciences from Goethe University, Frankfurt, Germany. After eleven years as a researcher and deputy chair in information systems in Germany, he joined HSG in 1996. His research areas include situational method engineering, enterprise architecture management, transformation management, healthcare management, and corporate controlling systems. He is vice editor-in-chief of *Business & Information Systems Engineering* (formerly *Wirtschaftsinformatik*), as well as a member of the editorial boards of *European Journal of Information Systems*, *Information Systems and e-Business Management,*” and *Enterprise Modelling and Information Systems Architectures.*

Communications of the Association for Information Systems

ISSN: 1529-3181