The State of Design Science Research within the BISE Community: An Empirical Investigation
Completed Research Paper

Joerg Leukel
Department of Information Systems 2
University of Hohenheim
70599 Stuttgart, Germany
joerg.leukel@uni-hohenheim.de

Marcus Mueller
Department of Information Systems 2
University of Hohenheim
70599 Stuttgart, Germany
marcus.mueller@uni-hohenheim.de

Vijayan Sugumaran
Department of Decision and Information Sciences
Oakland University
306 Elliott Hall, Rochester, MI 48309, USA
sugumara@oakland.edu

Abstract

The Business & Information Systems Engineering (BISE) community in the German-speaking countries has a long track record of publishing papers using design science research (DSR). However, the state of recent DSR within the BISE community is not well documented and the lessons learned can be useful for other communities. This paper investigates the use of DSR methodology by examining articles published in the BISE community's primary outlets. We focus on understanding the artifacts created, the foundations for building these artifacts, and the evaluation methods used. The results reveal a) a broad view of foundations for DSR by incorporating artifacts that are used in practice, b) the focus on the organization as the unit of analysis, c) a pluralism of research methods that cater to the timeliness of problems addressed, and d) low level of theoretical underpinnings, thus lacking in DSR rigor aspects.

Keywords: Design Science, evaluation methods and criteria, literature review, research methods/methodology, IT artifact

Introduction

Design science research (DSR) has become an important approach within information systems research. A particular community advocating DSR is the BISE community in the German-speaking countries (Austria, Germany, and Switzerland). BISE stands for Business & Information Systems Engineering, which is also the title of the community’s primary journal. Awareness of this community and its DSR proposition in the Management Information Systems (MIS) domain has increased since the publication of its memorandum on design-oriented research (Österle et al. 2011). According to the memorandum, the BISE community is deeply committed to DSR. The BISE community also foresees a potential to contribute to asserted problems within the MIS community such as lack of relevance of its research, flat enrolments, and decline of funding from industry (Buhl et al. 2012). Aside from this commitment, the literature does not adequately inform MIS researchers and practitioners about DSR adoption by the BISE community. Hence, a concerted effort is needed to review the current literature and assess the level of understanding of DSR and its adoption within the BISE community.
The MIS literature regularly reports on the state of research methods used, e.g., (Chen and Hirschheim 2004; Palvia et al. 2007). Similarly, past DSR in MIS has been analyzed based on several dimensions such as evaluation methods used (Peffers et al. 2012), artifact types created (Offermann et al. 2010), and most influential articles (Piirainen et al. 2010). These works contribute to a better understanding of DSR on a global scale. In contrast to this richness, examination of DSR within the BISE community is inadequate for three reasons: (1) none of the prior studies were specific to DSR but included other methodologies, (2) all studies used specific and divergent conceptualizations that make it difficult for the MIS community to comprehend and relate the findings, and (3) the most recent literature survey covered the years 2004 through 2007 (Becker et al. 2009), and thus could not reflect on the state of DSR in the BISE community, which experienced major changes over the past years. In addition, DSR is receiving increasing attention within the global MIS community, which has been bolstered by the publication of Hevner et al.’s research essay on DSR in MIS Quarterly (Hevner et al. 2004). Although the BISE memorandum contains no reference to Hevner et al.’s work, several key concepts were directly adopted (e.g., artifact classification, evaluation methods). Senior BISE scholars indeed admit that Hevner et al.’s work have had a significant impact on explicating the self-image of BISE and conducting DSR.

As of now, little attention has been given to assessing research rigor of the DSR conducted by the BISE community, which is in stark contrast to the behavioral research within MIS. This deficit makes it difficult to thoroughly validate claims and myths that surround the DSR in the BISE community. For assessing the rigor of DSR, we can turn to Hevner et al.’s (2004) information systems research framework and use three of its main elements of research rigor: (1) IT artifacts as research contributions, (2) foundations used from the knowledge base to build IT artifacts, and (3) evaluation methods. Prior surveys suggest that this framework is suitable for classifying pertinent research (Indulska and Recker 2008; Peffers et al. 2012; Venable 2010). This suitability may be due to its normative effects on authors, reviewers, and editors.

While there is a long tradition of conducting DSR within the BISE community, the results and the lessons learned have not been effectively communicated to the global MIS audience. A systematic analysis of the current DSR literature from BISE can help articulate the different approaches used, the quality and rigor of the research, the types of artifacts created, how they are used to solve specific problems, and what value they add. The objective of this paper is to undertake a comprehensive literature survey and content analysis of the current DSR literature and report on the state of DSR within the BISE community. Specifically, to gain a better understanding of the quality of the recent DSR being conducted by the BISE community, we seek to answer the following three research questions (RQ):

RQ1: What are the IT artifacts built by the BISE community?

RQ2: What are the foundations used for building IT artifacts by the BISE community?

RQ3: What are the evaluation methods used by the BISE community?

We describe the procedure and results of a literature survey and content analysis that we have conducted. We used two data sets: The first contained 80 articles published in Springer’s BISE journal from 2009 to 2013. The second data set contained 216 papers from the 2011 and 2013 proceedings of the International Conference on Wirtschaftsinformatik, which is the bi-annual flagship conference organized by the BISE community. We will refer to the proceedings as BISE proceedings, notwithstanding its original title. The two data sets have comprehensive coverage of outcomes produced by the BISE community and thus provide an appropriate foundation for answering the posited research questions.

The remainder of this paper is organized as follows. The next section provides the theoretical background to our research. Subsequently, we describe the research method and present the results. Then, we discuss the implications of our results for MIS researchers and conclude the paper.

**Theoretical Background**

This section introduces those elements of the information systems research (ISR) framework (Hevner et al. 2004) that will guide our survey, followed by a discussion of related findings from prior surveys on BISE.
**Information Systems Research Framework**

While there may be many alternate ways of structuring DSR, we argue that the ISR framework is useful for defining the nature of DSR as well as harmonizing the vocabulary for presenting DSR. The ISR framework combines behavioral-science and design-science paradigms to understand, execute, and evaluate research. For instance, the framework defines two research processes “develop/build” and “justify/evaluate” to cover behavioral research through developing and justifying theories as well as DSR through building and evaluating IT artifacts. Since our survey is concerned with DS, we will only be referring to elements that are relevant for DSR and using their DSR-specific terms. The ISR framework is comprised of three pillars – **Environment, IS Research, and Knowledge Base** – which are connected through relationships between: (1) Environment and IS Research and (2) IS Research and Knowledge Base. The former relationship describes research relevance, while the latter relationship depicts research rigor.

The **IS Research** pillar contains the processes for building and evaluating IT artifacts. The Knowledge Base pillar assembles all prior findings from ISR and its reference disciplines for grounding the build processes (denoted as Foundations) and guidelines for conducting the evaluation processes (denoted as Methodologies). Thus, rigor is the extent to which research uses applicable foundations and methodologies from the knowledge base to build and evaluate IT artifacts that ultimately advance the knowledge base. In summary, the conceptual framework of our survey is an adaptation of the ISR Framework by (1) restricting its coverage to DSR and (2) focusing on rigor. An overview of the survey framework is shown in Figure 1. Relevant parts of the original framework (IS Research and Knowledge Base pillars) appear in black, while the irrelevant part (Environment) is greyed out. The components of the IS Research and Knowledge Base pillars are briefly discussed below.

![Figure 1. Survey framework based on the ISR Framework (Hevner et al. 2004)](image)

**Build:** This process produces the IT artifact, which can be classified into four types (March and Smith 1995). **Constructs** provide the vocabulary to represent a phenomenon in the domain. Such artifacts may range from formalized notions, e.g., as available in conceptual modeling grammars, to informal higher order abstractions, which are more difficult to assess, e.g., psychometric constructs. **Models** are representations of domain phenomena built by using constructs. A model then contains a set of propositions or statements about the problem and solution space of the domain. **Methods** provide guidance for searching the solution space to solve problems. A method defines steps to be performed on a representation (e.g., model), of the solution space. Methods span from algorithms, which can be executed by computers, to informal guidelines, which are targeted at humans, e.g., system analysts, software engineers, and end users. Finally, **instantiations** are implementations of constructs, models or methods...
into working systems, i.e., software. An instantiation is a realization of another artifact in an environment for its intended purpose. Due to the many dependencies between these artifact types, a DSR project may build artifacts of more than one artifact type. Prior surveys have used this classification, e.g., (Samuel-Ojo et al. 2010), but its validity and completeness have also been the subject of inquiry and alternative classifications have been proposed (Offermann et al. 2010; Walls et al. 2004; Winter 2008).

**Foundations:** This part of the knowledge base provides prior research results that inform the build process as applicable knowledge. Foundations can be divided into: theories, frameworks and instruments contributed from behavioral research and IT artifacts produced by DSR. Building a new artifact must draw on current artifacts (e.g., by adding or improving particular properties) and behavioral theories (e.g., by justifying the design through known explanations of domain phenomena). Assessing the extent to which design-oriented researchers exploit the knowledge base is non-trivial but depends foremost on how well authors make design decisions transparent to the reader and relate their arguments to pertinent knowledge. Assuming that researchers appropriately report any characteristic of the build process, the paper serves as proxy for the process actually performed. This assessment is, however, made difficult by the sheer size and variety of foundations for which no widely accepted classifications exist so far.

**Evaluate:** This process demonstrates the utility of the IT artifact for solving the targeted problem. The ultimate goal is to determine if the artifact makes a contribution to the knowledge base by better solving the problem than with current artifacts and knowledge. Evaluation requires the definition of problem-relevant criteria and metrics, the selection of appropriate evaluation methods, and their rigorous execution. The ISR framework provides a guideline on design evaluation through a two-level classification of evaluation methods into five groups (as shown in figure 1) and twelve methods. The classification assembles a wide array of methods to reflect the diversity of IT artifacts and problem domains as well as epistemological stances. Additionally, multi-leveled evaluation that applies two or more evaluation methods to one artifact must be considered. As with artifact types, alternate categorizations have been proposed (Sonnenberg and vom Brocke 2012; Venable et al. 2012), which also include methods such as action research (as a particular form of case study), field experiment, and expert interview.

**Methodologies:** This part of the knowledge base provides data analysis techniques, formalisms, measures, and validation criteria for configuring an evaluation method that best fits the artifact and problem domain under study. While the framework defines a classification of evaluation methods (as discussed in the preceding paragraph), the knowledge base is much more comprehensive; it contains specific guidelines and techniques for planning and executing evaluation processes and interpreting their qualitative or quantitative results. The ability to identify the use of methodologies in a particular DSR depends on how authors present and externalize the evaluation process. This task may be negatively affected by the diverse backgrounds of researchers and lack of common practices, which then cause inconsistency in terminology. For overcoming these impediments, several classifications spanning foundations and methodologies have been proposed, e.g., (Vessey et al. 2002). Although these efforts provide immeasurable value to ISR, we conjecture that adoption by BISE researchers is lower compared to other communities.

**Prior BISE Surveys**

Table 1 provides a summary of prior surveys that were concerned, at varying degrees, with research rigor, e.g., by assessing scientific goals and research methods used. Although these surveys covered the entire BISE community, the dominance of the DSR approach as stated in the BISE memorandum (Österle et al. 2011), allows us to use them as surrogates for BISE’s take on DSR. Three surveys assessed articles published in the Wirtschaftsinformatik journal, which is the German ancestor of the BISE journal (the BISE journal is cover-to-cover identical to the former journal). Four surveys approached key BISE scholars for their beliefs and perceptions, and then interpreted the qualitative data.

Research rigor had received little attention within the BISE community until the mid-2000s (Buhl et al. 2012). The literature survey by Heinrich (2005) provides support for this assertion by reporting that only 11.0% of all research articles published in the community’s leading outlet contained an explicit statement of the research method used. Moreover, only one article belonged to meta-research by inquiring research methods. Heinrich then posited the provoking question whether BISE’s research may be attributed as science at all. An update of Heinrich’s study for the years 2004 through 2007 (Becker et al. 2009) detected some progress with 24.7% of papers articulating the research method used but still this period of four years saw only one meta-research article. The question by Heinrich may be further backed up by two
Delphi studies in which key BISE exponents participated (König et al. 1995a; Heinzl et al. 2001). Although both studies claimed to forecast the scientific objectives of BISE research in the next ten years, only the former study assessed the research paradigms and methods. Participants concluded that focusing on the problem-solving approach would be best to maintain BISE’s competitive position (74% agreement). A similar finding can be obtained from a survey of written autobiographies by 16 BISE scholars who belonged to the founding generation of BISE (Heinrich and Riedl 2013). These scholars provided personal reflections on the genesis and development of BISE. This survey confirmed the dominance of DSR as well as “a lack of awareness of the importance of theoretical research” (p. 40). Similarly, the literature survey by Wilde and Hess (2007) determined the share of DSR articles at 71% for the years 1996 through 2006.

<table>
<thead>
<tr>
<th>Survey</th>
<th>Unit of observation</th>
<th>Items</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>König et al. (1995a, 1995b)</td>
<td>BISE scholars (23) and practitioners (7)</td>
<td>30</td>
<td>Delphi method</td>
</tr>
<tr>
<td>Heinzl et al. (2001)</td>
<td>BISE scholars (26) and practitioners (4)</td>
<td>30</td>
<td>Delphi method</td>
</tr>
<tr>
<td>Frank et al. (2008)</td>
<td>BISE scholars</td>
<td>8</td>
<td>Structured interview</td>
</tr>
<tr>
<td>Heinrich and Riedl (2013)</td>
<td>BISE scholars</td>
<td>16</td>
<td>Written autobiography</td>
</tr>
</tbody>
</table>

Table 1. Prior surveys of research in the BISE community

With regard to IT artifacts built (RQ1), no survey assessed particular artifact types as defined in the ISR framework. However, three surveys accent the role of software prototypes, thus instantiations, as the prime outcome of design-oriented research (Frank et al. 2008; Heinrich and Riedl 2013; König et al. 1995a). This importance is also being reflected in the consideration of prototyping as a research method on its own, which entails both the build and evaluate processes in DSR (Heinrich and Riedl 2013).

Foundations informing the design of research artifacts (RQ2) have not been explicitly studied, though some findings can be drawn from the debate on the role of reference disciplines for BISE. The forecast by the first expert panel ranked organization theory/studies within business administration as the most important reference for BISE research (König et al. 1995a). In a similar vein, senior scholars agreed that BISE had its origin in business administration (denoted as a “mother discipline”) (Heinrich and Riedl 2013). While the second Delphi study did not revisit this question (Heinzl et al. 2001), the experts deemed “foundations of BISE” and “interfaces of BISE to other disciplines” as the third-least and least relevant topic of research (within a list of 14 topics).

When discussing the use of evaluation methods (RQ3), we must be aware of differences in the conceptualization of methods within BISE as compared to the one in the ISR framework. These differences were quite big in the past, as can be seen from the first Delphi study (König et al. 1995b). This study assigned empirical methods such as case study, field study and interview solely to behavioral research and theory development; hence, these methods were regarded as not available for evaluating the utility of artifacts (unlike in the ISR framework). This very unusual limitation, however, must not be attributed to the study’s design because such a misconception could have been rectified by the experts in the course of performing the Delphi study. The prime methods posited for the problem-solving approach were “development and test of prototypes” and simulation. Thus, it can be stated, that the early BISE community understood methods in the sense of development methods rather than research methods (Heinrich and Riedl 2013).

Nevertheless, evaluation methods have been touched in all three literature surveys. There are two difficulties in comparing their findings: First, these surveys covered design-oriented and behavioral research by assessing research methods, thus their results cannot be directly related to DSR. Second, each survey used a specific conceptualization of methods that differed more or less from each other. Still, the surveys might provide some indications of preferences within BISE.

Interpreting the survey by Heinrich (2005) suffers from the small sample size that was available for research methods (as discussed above, only 11.0% of all articles reported the method used, thus N=59). The same constraint holds for the succeeding survey by Becker et al. (2009) with N=24. Any comparison...
is hardly possible due to conflicting terminology and missing definitions of each method. For instance, Heinrich subsumed various DSR evaluation methods under “modeling and implementation (construction of prototypes), partly with testing (e.g., simulation)” (p. 108), which was assigned to 13.6% of the articles. Nevertheless, both surveys found varying use of methods, which we denote in their original terminology: survey (50.9% vs. 37.5%), simulation (n/a vs. 25.0%), laboratory experiment (16.9% vs. n/a), and case study (n/a vs. 20.8%). The survey by Becker et al. reports “design science” (8.3%) as a method separately. This oddity might be explained by the surveyors directly taking the author’s description of the method used but not relating it to a sound taxonomy of disjoint methods.

The survey by Wilde and Hess (2007) was guided by a broad set of 14 research methods. Each method was classified into a portfolio along two dimensions, namely “degree of formalization” (qualitative vs. quantitative) and “research paradigm” (behavioral vs. design-oriented). Similar to the two former surveys, laboratory experiments and case studies were assigned to behavioral research but not considered for evaluating IT artifacts. Wilde and Hess provide definitions of each method and point to exemplar articles in Wirtschaftsinformatik (if available). We are able to interpret some results as follows: the most used methods were informed argument and scenario (45%, denoted as argumentative and semi-formal deduction), case study (16%), and prototyping (13%). Use of controlled experiment, simulation and action research was very low (each less than 2%), whereas field experiment was not found at all. It might be surprising to spot descriptive evaluation methods in the first place, considering that they may represent the ‘last resort’ if other methods are not feasible. However, Wilde and Hess also noted a decreasing share of descriptive evaluation, with 22% for the last three years of the sample (2004 through 2006).

In light of the surveys on BISE (discussed above) and the knowledge they provide to understand DSR adoption within this community, our research design aims at mitigating three particular shortcomings. First, our survey is a comprehensive endeavor that exclusively focuses on design-oriented research and thus is more appropriate to assess DSR adoption, whereas prior surveys were not specific to DSR and included behavioral research. Second, by grounding our survey on Hevner et al.’s ISR framework, we strive for enhancing the abilities of global MIS communities to comprehend and appreciate DSR adoption by the BISE community due to exerting a widely known conceptualization of design science. We acknowledge, however, that this approach may conceal certain properties that the BISE community itself regards as a characteristic, e.g., prototyping as a research method. Third, by covering the full lifetime of the BISE journal and its latest volume of 2013, the survey allows to depict a contemporary picture of DSR adoption and identify changes that have occurred in the on-going academic debate on the design science paradigm.

Method

To answer the research questions, we conducted a literature survey on two data sets representing major outcomes of the BISE community:

- The initial journal data set included all articles published in volume 1 through 5 of the BISE journal (from 2009 to 2013), thus, since the inception of this English journal that complemented the German Wirtschaftsinformatik journal. The BISE journal is sponsored by the MIS sections within the German Informatics Association (GI) and the German Academic Association for Business Research (VHB), and thus has a similar role as the Journal of the Association for Information Systems. The journal is the result of a strategic realignment of the main publication outlets within the BISE community. As a result, five out of seven journal departments are co-edited by non-BISE scholars and international readership has increased. We only considered papers listed in the Research Paper category but not State-of-the-Art, Research Note, and Catchword. We also excluded the inaugural issue, which contained seven visionary papers and eight seminal papers from the past but no new technical contributions. Thus, the initial data set had 80 articles. All articles were available in electronic format from SpringerLink via AIS eLibrary.

- The initial proceedings data set included all papers published in the BISE 2011 and 2013 proceedings. These proceedings only contained completed-research papers (N=216). All papers were available in electronic format from AIS eLibrary.

The analysis procedure consisted of the following four main tasks, which were carried out from December 2013 to August 2014.
1. **Identification of DSR papers:** To identify those papers that build and evaluate an IT artifact, the first two authors independently examined each journal paper and marked whether it fulfills the criterion. There was a high inter-rater agreement between the two, indicated by Cohen’s kappa at k=.895. Conflicting ratings were resolved by discussing each paper in detail. In total, 48 out of 80 papers were classified under DSR. For the conference papers, the same procedure was followed by the first author and one PhD student. Inter-rater agreement was high with k=.840. The final data set included 97 conference papers (of which 59.8% were in English and 40.2% in German).

2. **Pilot study with coders:** To reduce research bias, the content analysis of journal papers was performed by two PhD students. For instructing the coders (the two PhD students), we first explained the codebook and then held three training sessions. The codebook described the criteria for the four artifact types and twelve evaluation methods as discussed in the preceding section. As for the foundations criterion, coders were required to identify from each paper those terms used by the authors to denote the relevant state-of-the-art (which can often be found in a section on “related work” or “theoretical background” in DSR papers) as well as theories that the authors referred to, if any. The objective of the training sessions was to validate the codebook for a sample taken from the proceedings data set. The students independently coded five papers, which were then discussed with the instructors to find agreement on each criterion. This procedure was repeated two more times. Altogether 15 papers were coded and discussed, so that we could expect a sufficient level of inter-coder agreement in the main study.

3. **Content analysis of journal papers:** This step was organized similar to the pilot study, with the coders independently analyzing each paper and coding the data. This was followed by a discussion of each paper under the supervision of the instructors. Because of the variety of evaluation methods available in the DSR literature, the codebook allowed to suggest new categories apart from the twelve methods found in the DSR framework. In this case, the coder was asked to link the proposed method to one of the five categories. The procedure thus had an explorative element as follows: The main study commenced with coding of the articles in volume 5 (2013) of the journal data set. The codes were discussed with the instructors to resolve conflicts. Then, the coders turned to the articles in the preceding volume (volume 4), and so on. For agreeing on the terms that appropriately represent the foundations used, the group discussion involved an additional task. The group removed slight variations and synonyms to uncover common foundations across the authors’ terminology (e.g., “decision support” and “management support systems”).

For volume 4 (2012), both coders independently reported the need for adding at least one evaluation method. These were found in five papers under the following titles: “expert interview”, “focus group”, “expert survey”, “expert discussion and interview”, and “expert focus group”. These suggestions were discussed by the group and it was decided to add one collective method. All five reported evaluation processes have one thing in common; the artifact was assessed by a group of domain experts in an experimental environment (no real business environment), however, the treatment was given to all subjects. Thus, unlike in a controlled experiment, no control group existed but only the treatment group. This experimental variant was added to the taxonomy as “expert evaluation” and became available for the next iteration of coding. The average initial agreement for all volumes was moderate to substantial for artifact types (72.9%) and evaluations methods (60.4%) for artifact types but perfect for foundations (100.0%). The group discussions lasted about one hour per volume and finally led to mutual agreement between coders and instructors.

4. **Content analysis of conference papers:** The main study concluded with the set of 97 conference papers. The first two authors independently coded each DSR paper in the 2013 volume in a single iteration. Initial agreement was substantial for artifact types (76.1%) and evaluation methods (69.6%) and perfect for knowledge base (100.0%). Then, the DSR papers in the 2011 volume were independently coded by the first author and one PhD student. Similarly, agreement was substantial or high (84.3% for artifact types, 70.6% for evaluation methods, and 82.4% for foundations), and conflicting codes were resolved afterwards. The discussion of these codes took about four hours.

**Results**

DSR papers account for 60.0% of the initial journal data set (N=80). As shown in Figure 2, this percentage has increased since 2011, which represented the year in which design-oriented research was
even outnumbered by other research. We assessed which of all the DSR papers (N=48) contained a reference to Hevner et al.’s work (2004). While the share varied during the five years, almost all of the most recent DSR papers contained a reference. Many set out their research approach by referring to elements of the ISR framework. With respect to the proceedings, the percentage of conference papers that provide such a reference is lower (24.7% for the years 2011 and 2013). This may be attributed to the page limit giving lesser space for explicating the research approach in detail.

Artifact types: The most frequently used artifact type is method with on average 66.7% (journal) and 61.9% (proceedings). Next is model (29.2% and 25.8%), then instantiation (12.5% and 10.3%), and finally construct (10.4% and 5.2%). In the last two years, the percentage of model artifacts in the BISE journal has increased at the expense of method artifacts (as shown in Figure 3). Concerning the number of artifacts types, most papers proposed artifacts of one type (83.3% and 96.9%), very few included two types (16.7% and 3.1%), and none more than two types. The most frequent combinations were construct/method, and model/method.

**Figure 2. Percentage of DSR papers (bars, N=80) and references to Hevner et al. (line, N=48) in the BISE journal**

**Figure 3. Percentage of IT artifact types used in the BISE journal (N=48)**

Foundations: For analyzing the foundations, we performed several steps. First, we reviewed the list of terms that was available for each paper. We classified these terms into the two categories, namely artifact and theory. These categories were regarded as distinct but complementary as defined in the ISR framework. That is, theory had a rather narrow meaning, by restricting it to explanatory theories as used in the social sciences. The rationale for this procedure was to be able to detect to what extent BISE papers draw on theories that are relevant to behavioral research in MIS. Then, we manually harmonized the level of abstraction for related terms. For instance, “process reference models” was merged with the more abstract term “reference models” and “requirements specification” with “requirements engineering”. This step was aimed at condensing the broad set of terms into larger entities.

The journal data set of 48 papers provided 108 initial terms. Eight papers had only one term, 20 papers had two terms, 16 papers had three terms, and four papers had four terms. From the 108 terms, we identified the following four theories: “collective action theory”, “knowing organization theory”, “resource-based view (of the firm)”, and “systems theory”, which each appeared once. By aggregating related terms, we finally arrived at 77 terms. Still, these terms represent a wide array of foundations since 84.4% of these terms occurred only in one paper each. Next, we list all the terms with frequency greater than two: “business process management” (10), “requirements engineering” (6), “service-oriented architecture” (5), “cloud computing” (3), “reference models” (3), and “semantic technology” (3).
The foundations results for the proceedings data set are as follows: 97 papers yielded 108 initial terms. Three terms denote theories (“mechanism design”, “game theory”, and “auction theory”). The aggregation step reduced the number of terms to 81, which still represent a great diversity in foundations. This finding is also reflected in the short list of terms with frequency greater than two as follows: “business process management” (12), “service-oriented architecture” (7), and “enterprise architecture” (3).

Evaluation methods: The data presented in Table 2 indicates significant changes in evaluation practices over the past five years.

<table>
<thead>
<tr>
<th>Category of Evaluation Methods</th>
<th>Specific Evaluation Method</th>
<th>Outlet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Journal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009 (N=13)</td>
</tr>
<tr>
<td>Observational</td>
<td>Case study</td>
<td>30.8%</td>
</tr>
<tr>
<td></td>
<td>Field study</td>
<td>-</td>
</tr>
<tr>
<td>Analytical</td>
<td>Static analysis</td>
<td>7.7%</td>
</tr>
<tr>
<td></td>
<td>Architecture analysis</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Optimization</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dynamic analysis</td>
<td>-</td>
</tr>
<tr>
<td>Experimental</td>
<td>Controlled experiment</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Simulation</td>
<td>15.4%</td>
</tr>
<tr>
<td></td>
<td>Expert evaluation</td>
<td>7.7%</td>
</tr>
<tr>
<td>Testing</td>
<td>Functional (black box)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Structural (white box)</td>
<td>-</td>
</tr>
<tr>
<td>Descriptive</td>
<td>Informed argument</td>
<td>15.4%</td>
</tr>
<tr>
<td></td>
<td>Scenario</td>
<td>46.2%</td>
</tr>
<tr>
<td>No evaluation</td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2. Percentage of Evaluation Methods Used

In 2009 and 2010, descriptive methods in journal papers accounted for 61.6% and 40.0% respectively, but this has dropped to 11.1% in 2013. At the same time, experimental methods were more often used, with preference on simulation and controlled experiment. In parallel to the decrease of single case studies, expert evaluation decreased sharply in 2013. Another finding is that all six analytical and testing methods are marginal. The results for the proceedings data set largely concur with those of the journal data set. Considering the lower acceptance criteria of proceedings, methods that require extensive data collection and analysis were seldom found. The much lower barrier may also explain why almost a fifth of all conference papers lacked any form of evaluation, contrary to each journal paper reporting an evaluation. In addition, we checked the number of evaluation methods per paper. We found one evaluation method in 89.6% of the journal papers, and 81.4% of the proceedings papers, while two methods were reported in 10.4% and 3.1%, respectively.

In the next step of our data analysis, we studied the dependence of evaluation methods on the type of artifact that was evaluated (Table 3). The ability to interpret this data for the journal may be limited due to the changes that have occurred over time as shown in Table 2. However, we can point to three observations. First, the evaluation of constructs, models and methods almost evenly splits between three methods/categories, i.e., case study, experimental, and descriptive methods. Second, we did not observe any abnormal combination such as arguing for the usefulness of proposed instantiations by providing informed arguments or constructing scenarios. On the contrary, most dependencies make a lot of sense, e.g., instantiations were coupled with experimental methods (66.7%). Third, models were the primary object of expert evaluation; in all four cases, these models were reference models (data, process, quality) and the experts were asked whether they perceive the proposed model as useful.
The proceedings data in the right hand side columns of Table 3 indicate similar dependencies, although at the lower end of evaluation methods. Hence, experimental methods focus on simulation rather than involving subjects that interact with the artifact in some form, and models have been either applied once in an organization, assessed in a descriptive form, or not evaluated at all.

<table>
<thead>
<tr>
<th>Category of Evaluation Methods</th>
<th>Specific Evaluation Method</th>
<th>Outlet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observational</td>
<td>Case study</td>
<td>40.0%</td>
</tr>
<tr>
<td></td>
<td>Field study</td>
<td>-</td>
</tr>
<tr>
<td>Analytical</td>
<td>Static analysis</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Architecture analysis</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Optimization</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dynamic analysis</td>
<td>-</td>
</tr>
<tr>
<td>Experimental</td>
<td>Controlled experiment</td>
<td>20.0%</td>
</tr>
<tr>
<td></td>
<td>Simulation</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Expert evaluation</td>
<td>-</td>
</tr>
<tr>
<td>Testing</td>
<td>Functional (black box)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Structural (white box)</td>
<td>-</td>
</tr>
<tr>
<td>Descriptive</td>
<td>Informed argument</td>
<td>20.0%</td>
</tr>
<tr>
<td></td>
<td>Scenario</td>
<td>40.0%</td>
</tr>
<tr>
<td></td>
<td>No evaluation</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3. Evaluation Methods per Artifact Type

Discussion

The goal of this work has been to better understand the quality of the recent DSR being conducted by the BISE community and inform the global MIS community about the dominating practices of BISE research. The results provide insights into these practices which we will discuss at four levels as follows: (1) outlets for DSR, (2) IT artifacts and foundations, (3) evaluation methods, and (4) lessons learned.

**DSR Outlets**

The first result, though not surprising, is that the two BISE outlets studied have the highest receptivity for DSR compared to all other MIS outlets. The share of design-oriented articles for the journal is 60.0% and for the proceedings is 42.1%, respectively. Although all major MIS outlets nowadays are open to DSR and have included this approach into their editorial statements (Baskerville et al. 2011), the actual percentage of DSR is much lower on the global scale. Exact data is not available but recent years suggest a share of 5% to 10% in the top MIS journals. Besides the number of articles, perceptions of the members of the community are important to appraise receptivity. A recent survey among 57 active design-oriented researchers investigated the perceived receptivity of 60 journals (VanderMeer and Tremblay 2013). Although the BISE journal ranked only 12th (3.6 on a 5-point scale), from the superior journals all but three were Computer Science outlets. Acknowledging that MIS researchers perceive the BISE journal as being receptive for DSR, the study also found that the respondents did not perceive the journal’s impact on the same level.

The share of DSR papers in the BISE proceedings also exceeds all other MIS conferences. The data on MIS conferences is richer than for journals but partly contradictory. In the most comprehensive survey of more than 7,500 papers published from 1999 to 2008, Olbrich (2009) reported quite high percentages for ICIS (25%), AMCIS (9%), and ECIS (35%). However, these findings are inconsistent with the perceptions...
articulated by eminent supporters of DSR as well as the survey by Indulska and Recker (2008), which covered the more recent period from 2005 to 2007. From a sample of 3,284 papers contained in the proceedings of ACIS, AMCIS, ECIS, ICIS and PACIS, they found only 2.5% DSR papers. Indulska and Recker also noted an over-proportional share of DSR contributions from European authors, in particular, authors from German-speaking countries.

**IT Artifacts and Foundations**

The main research outputs of the BISE community are methods, which account for about two thirds of all artifacts (though this has dropped in the past two years for the journal). A comparison with findings from the only existing survey that also considered artifact types (Samuel-Ojo et al. 2008) suggests significant differences. This survey examined all 92 papers that were presented at DESRIST conferences between 2006 and 2009. DESRIST is the first conference series in MIS that is exclusively dedicated to DSR. The survey reported models in the first place (40.2%), followed by methods at 23.9%. In the design-science paradigm, methods are the ultimate means for problem solving, since they provide specific guidance on how to solve problems and rely on representations such as models. Following this view, the dominance of methods in DSR by the BISE community suggests a closer orientation towards the problem solving criteria and the applicability of these artifacts than DESRIST.

In carrying out the content analysis, we observed not only a wide range of methods but also noted an emphasis on methods that assist in strategic decision making by the IT and business executives rather than operational decisions. The proposed methods are more frequently targeted at the organizational level than the individual level (difference in unit of analysis). In many cases, these methods were developed in close cooperation with domain experts, underwent several iterations of build-evaluate processes, and were evaluated with respect to requirements posited by these domain experts. Actually, specific research methods have been proposed by BISE researchers for conducting this form of research. Of particular significance is the so called “consortium research” approach (Österle and Otto 2010), which is intended to support the researcher in accessing and capturing knowledge from practitioners. While this method has been applied and refined for more than 20 years, it was also proposed as a method artifact of DSR, evaluated by means of a longitudinal field survey, and presented by referring to Hevner et al.’s guidelines on design-oriented research.

Consortium research stresses the importance of knowledge from artifacts that are already in use, and that researchers must assess this knowledge to be able to make any meaningful contribution to solving practical problems. This view of knowledge has also found its way into the BISE memorandum by stating that the body of knowledge is constituted by the literature but “to a much larger extent [...] by the experiences and knowledge accumulated in business” (Österle et al. 2011, p. 8). Our content analysis provides evidence for the broad foundations that BISE researchers use when building artifacts. On the other hand, explanatory theories played a marginal role and were only found in a handful of journal and conference papers. Particular MIS theories were not observed at all. This interpretation of foundations is much more extensive than the DSR conception of Hevner et al. In particular, the ISR framework indicates that foundations are provided by “IS research and results from reference disciplines” (Hevner et al. 2004, p. 80). Thus, has theory no place in BISE research? If we rely on the memorandum, then theory serves the underlying design decisions. Yet, our survey found little evidence for deriving design elements from existing theories. This assessment, however, must be seen in the general context of the range and role of theory in DSR.

In an ideal world, theoretical propositions guide design research and justify elements of the design. However, a field study of 68 cross-discipline DSR scholars identified a great variety of theory use, ranging from theory as being implicit or tacit, and theory as ontology that provides constructs for representing domain phenomena, and finally theory as concrete descriptions for building artifacts (Haynes and Carroll 2010). In the BISE samples, we found a similar array of theory use. The current minimal role of theory may be due to limited power of extant theories to explain and predict new domain phenomena that arise from advancements in IT and business practices (Lee 2010). BISE researchers are more exposed to these advancements than MIS researchers (Steininger et al. 2009). The foundations that we identified from the samples were referring to IT artifacts for topical problems of urgency for practitioners.
Evaluation Methods

Of particular importance are evaluation methods that allow BISE researchers to engage with practitioners. Assuming that methods follow topics and the problem that the IT artifacts address, prototyping has been put forward as a specific method of utmost importance. The role of prototyping in engaged scholarship was emphasized in all recent surveys on the BISE community (e.g., Becker et al. 2009; Heinrich 2005; Heinrich and Riedl 2013; Wilde and Hess 2007) as well as in the BISE memorandum. With respect to the more commonly acknowledged definition of evaluation processes in Hevner et al.’s essay, a terminological problem surrounds prototyping as a specific research method. The problem manifests in the definition used in the study by Wilde and Hess, which reads as follows: “development and evaluation of a preliminary version of an application system.” (2010, p. 282, translated from German). This seems to suggest that prototyping does not prescribe how to evaluate the application and what specific method to use for conducting the evaluation. For instance, the prototype system could be implemented in an organization (case study), used in a controlled environment (experiment) or tested with artificial data (simulation).

The mismatch of the memorandum’s conceptualization with the ISR framework also affected our content analysis, which had to relate the terminology found in BISE papers to the taxonomy of evaluation methods in the ISR framework. While this could be resolved, to a large extent, by the procedures used in the survey method, these differences might undermine the ability of the global MIS community to directly interpret outcomes from BISE research. The survey results, however, suggest that BISE researchers have started to align their terminology with the global definition of methods. As for the relevance of particular evaluation methods, we noticed a preference for experimental methods over descriptive evaluation, which prevailed in the past. This observation corroborates the recent suggestions within the BISE community to pay more attention to rigor in conducting the evaluation process (Buhl et al., 2012).

Preference is given to simulation and controlled experiment but our data is still ambiguous about the role of expert evaluations. The increasing importance of controlled experiments involving subjects may in the long run affect the problems addressed by the individual as the unit of analysis instead of the organization. In earlier times, many BISE researchers were reluctant to examine artifacts first in a laboratory setting but favored a real-world setting of an organization (Heinrich and Riedl 2013). This stance was largely due to the nature and relevance of problems addressed and the institutional context of BISE research. Our findings suggest that BISE researchers recently tend to put more emphasis on the control of confounding factors for internal validity than in the past, which was guided by external validity.

Lessons Learned

Our analysis of the literature from the BISE community reveals that while DSR has been widely adopted, the approaches used vary quite a bit because the research has been driven by specific industry needs and different problems being solved. Most of the research has also been centered on facilitating strategic decision making at the organizational level. One of the lessons learned is that this “consortium approach” to DSR did not take into account the typical controls and explanatory theories that are an integral part of DSR and hence the research may not be considered rigorous by the main stream MIS community.

The understanding of what theoretical propositions are relevant to DSR and how to translate theory into activities by the researcher is still limited. The DSR literature does not sufficiently inform the general procedures as well as provide many instances of applying successful theory to design. A rich body of frameworks, principles, procedures, and guidelines for theory-led design has emerged (Carlsson et al. 2011; Gleasure et al. 2012; Kuechler and Vaishnavi 2012). While the more recent papers from the two BISE samples contained several adoptions of this body of knowledge, still only a handful of articles provided thorough reasoning about the theory-component of their design but rather framed their research approach on a higher level of abstraction.

Another lesson learned is that the DSR researchers haven’t been consistent in the use of terminology and hence their works run the risk of being misinterpreted by the global DSR community. Very often, the practice-driven nature of DSR is evident from the outset of the papers and the way they articulated the research design. While most DSR papers claim to follow the DSR methodology by referring to Hevner et al.’s article, the actual impact of their specific ISR framework is still low. That is, few articles subscribe to the DSR terminology and then use its concepts correctly. We found such deviations for artifact types,
evaluation methods, and foundations. For instance, some BISE researchers mixed up the terminology into custom descriptions of their evaluation approach such as “case-study-based simulation” and “illustrative case example”, or regarded controlled experiments as an ideal that couldn’t effectively be achieved by research (stated in a BISE journal article of 2010). Quite often, DSR appears to serve as a “label” used for “selling” the research but doesn’t consistently materialize throughout the presented research.

Sufficient care has to be taken to ensure that the basic tenets of the DSR methodology are strictly followed in order to gain credibility outside the BISE community. While the DSR research within the BISE community has led to some ingenious solutions to some complex problems, the lack of rigor in some cases limits them from being generalizable and broadly applicable. Based on these observations, our study acts as a magnifying glass for various facets of general DSR in MIS. Unlike behavioral research, and in particular quantitative positivist research, very few articles have a recognizable publication formula. The diversity is not limited to the range of problems addressed but also concerns the types of knowledge contributions and the means used for validating and presenting the research (Dwivedi et al. 2014). We hope that recent debates within the DSR communities to clarify the knowledge contributions, research processes, and presentation, such as the DSR communication schema proposed by Gregor and Hevner (2013), will result in more effective articulation of DSR efforts, and thus improve DSR appreciation within the MIS community.

**Conclusion**

This paper has investigated the use of design science research methodology within the BISE community by adopting Hevner et al.’s ISR framework (2004). Specifically, we focused on understanding the types of artifacts created, the foundations used to build these artifacts, and the evaluation methods used. We examined two sets of journal and conference papers, which included 48 and 97 DSR papers, respectively. While design-oriented research is predominant within the BISE community, the applications tended to have a managerial focus and the development of the artifacts don’t necessarily have well-articulated theoretical underpinnings. In other words, the rigor aspect of DSR is somewhat lacking. On the other hand, the contributions of DSR emanating from the BISE community are as follows: A broader view of the foundations for DSR appears which incorporates artifacts that are not described in extant literature but used in practice and provides means for accessing and appraising this part of the knowledge base. The focus on the organization as the unit of analysis may complement findings from extensive MIS studies on the individual and group level, and thus also contribute to theory. A pluralism of research methods may better cater to the timeliness of problems addressed and allow the researcher to engage with practice. There is still a need for further aligning these specific DSR contributions with terminology and developments in the global DSR and MIS communities (e.g., conceptualization of research methods, adoption of practices for conducting experimental research, instrumentation of theoretical constructs, and data analysis).

**Acknowledgements**

We like to thank Marvin Hubl, Johannes Merkert, and Martin Riekert for their assistance in the document analysis.

**References**


