Design Science Research Post Hevner et al.: Criteria, Standards, Guidelines, and Expectations

John R. Venable

School of Information Systems, Curtin University, G.P.O. Box U1987, Perth WA 6845, Australia
j.venable@curtin.edu.au

Abstract. There is ongoing debate about how the quality (rigour and relevance) of Design Science Research (DSR) should be judged. This research investigates the state of the debate by surveying the opinions of IS scholars who write, review, edit, and publish DSR papers. The survey respondents rated the relative importance of the seven guidelines (often used as evaluation criteria) laid out in Hevner et al. (2004) [6], more specific criteria about the evaluation activity in DSR, criteria concerning IS Design Theories, and miscellaneous other criteria, and made general open-ended comments. The findings indicate a lack of consensus, with much variability in ratings. The Hevner et al. [6] guidelines are largely endorsed, but caution is also raised to apply them less mechanistically than at present. Some criteria/guidelines are seen to be less important at earlier stages of research. Caution is also urged not to expect single papers to fit all criteria/guidelines.

Keywords: Design Science Research, Research Method, Research Standards, Evaluation, IS Design Theory.

1 Introduction

It has now been more than five years since the publication of Hevner, March, Park, and Ram in MIS Quarterly [6], which set the de facto standard for the conduct and evaluation of Design Science Research (DSR). Since then, there is healthy debate about how the quality (rigour and relevance) of DSR should be judged and what quality goals DSR should attempt to achieve. (e.g. [1, 12, 2, 3, 15, 16], panel discussions at DESRIST 2007, ECIS 2007 and ACIS 2007, and an entire conference (Information Systems Foundations: Answering the Unanswered Questions about Design Research, 2008). Still, in the author's experience from continuing conversations on these topics, open questions remain concerning the practicality and appropriate application of the Hevner et al. [6] guidelines, the role and form of IS Design Theories [18, 5], what (if any) more detailed standards or guidance concerning the evaluation activity in DSR (Hevner et al. 2004 Guideline #3) are appropriate, and whether various other goals and criteria should also be applied when evaluating DSR.

The objective of the research reported in the current paper is to develop a better understanding of the current state of the IS field's dialog and argument about appropriate criteria, standards, guidelines, and expectations for DSR. It seeks to answer the following research questions:

R. Winter, J.L. Zhao, and S. Aier (Eds.): DESRIST 2010, LNCS 6105, pp. 109–123, 2010.

- What is the level of consensus within the IS DSR field concerning criteria, standards, guidelines, and expectations for how IS DSR should be conducted and reported?
- What other views are there besides the received view of Hevner et al. [6] concerning how DSR should be conducted and reported?

This research investigates the opinions of members of the scholars who are engaged in that dialog through the processes of writing, reviewing, editing, and publishing DSR papers using a survey. The survey asks questions concerning the relative importance of the seven guidelines laid out in Hevner et al. [6], more specific criteria for the evaluation activity in DSR, criteria for the development and reporting of IS Design Theories [18, 5] in DSR, and miscellaneous other criteria concerning the process and output of DSR.

The paper is organized as follows. First, literature relating to criteria, standards, and expectations in IS DSR is briefly reviewed. In section 3, the survey method and sample are described. Following that, the results of the quantitative data analysis are described, as well as reporting on other issues in DSR raised by the survey respondents. Finally, the paper identifies key issues remaining to be resolved and makes recommendations for further research.

2 Literature Review

Despite the relative new recognition of its status in the IS field, there is an extensive literature both before and after Hevner et al. [6] about IS DSR, much of which is relevant to the concerns of criteria, standards, and expectations for IS DSR.

2.1 Pre Hevner et al. (2004)

While Hevner et al. [6] set the de facto benchmark for DSR, important earlier work helped set the stage. Aspects of criteria, guidelines, and areas of expectation are highlighted briefly here.

Nunamaker, et al. [11] wrote a seminal paper in design science, which focused on justifying system development (a subset of design science) as an IS research method. Their "Multimethodological Approach to IS Research" (figure 2, p. 94) included a theory building activity, which addressed "development of new ideas and concepts, and construction of conceptual frameworks, new methods, or models" (p. 94) as well as theories. Their five stage process (with backtracking) included "construct a conceptual framework" as a first step in which the researcher should "(a) declare the 'truth'", "(b) formulate a concept (i.e. a framework)", "(c) construct a method", and "(d) develop a theory" (p. 99) and an evaluation step in which one should "Develop new theories/models based on the observation and experimentation of the system's usage" (p. 98).

Walls et al. [18] identified theory as a desirable output of Design Science Research. They proposed that an Information Systems Design Theory (ISDT) should "be a prescriptive theory which integrates normative and descriptive theories into design paths intended to produce more effective information systems." (p. 36). They proposed that an ISDT would have seven components. Four components concern the design product,

including (1) meta-requirements, (2) meta-design, (3) kernel theories, and (4) testable design product hypotheses. Three components concern the design process, including (5) design method, (6) kernel theories, and (7) testable design process hypotheses.

March and Smith [9] made two main contributions. First, they identified two main DSR processes: build and evaluate. Second, they identified four kinds of design artifacts (outputs of DSR): constructs, models, methods, and instantiations. In contrast to Nunamaker et al. [11] and Walls et al. [18], March and Smith [9] implicitly rejected the idea of design theories. "Rather than producing general theoretical knowledge, design scientists produce and apply knowledge of tasks or situations in order to create effective artifacts." (p. 253)

Venable and Travis [17] built on the work of Nunamaker et al. [11], emphasising the key role of theory building, extending Nunamaker et al.'s notion of a Computer-Based Information System as the designed artifact to include IS Development Methods, Tools, and Techniques as relevant designed artifacts , and refining the Multimethodological IS Research Framework [11] to substitute "In Situ Investigation" for "observation" and include Action Research as a method for In Situ Investigation.

Markus et al. [10] used the ideas of the ISDT of Walls et al. [18] but developed more "layman's" terminology for the concepts of meta-requirements and meta-design.

Rossi and Sein ([13] in acknowledged collaboration with Purao) added "better theories" (p. 5) to the four design artifacts in March and Smith [9] and "theorise" (p. 6) as a step in Design (Science) Research. During evaluation, Rossi and Sein (with Purao, [13]) propose both internal and external criteria. Among the internal criteria is "Match between the artifact and the 'abstract idea'. How well does the artifact embody the abstract idea that is being researched?" (p. 8). This concerns the constructs or concepts and how faithfully they are implemented in the models, methods, and instantiations. Among the external criteria is "Advancement of design theory: Is the abstracted idea generalisable to other contexts or at least advance our understanding of other design contexts?" (p. 9). This proposes that goals for good design theory would include generalisability of the artefacts and the utility of the design artefacts in other problem contexts.

2.2 Hevner, March, Park, and Ram (Hevner et al. 2004)

As noted in the introduction, the paper by Hevner, March, Park, and Ram [6] represents the received view of the IS field's conception of DSR. The publication has become the most-cited work in *MIS Quarterly*, the highest rated journal in the IS field, and is very influential. The authors draw on the earlier work of March and Smith [9] and others to develop an overall IS Research Framework as well as guidelines for the conduct and reporting of Design Science research.

Among other things, Hevner et al. [6] proposed seven guidelines for Design Science in IS Research, which can be summarised as follows:

1. Design as an Artifact – An identifiable and viable design artifact, as in March and Smith [9], must be produced.
2. Problem Relevance – The design must address a relevant and important problem.
3. Design Evaluation – The utility, quality, and efficacy of the design artifact must be rigourously evaluated.

4. Research Contributions – The contribution must be clear and verifiable. Contributions are seen to arise out of the novelty, generality, and significance of the designed artifact. Contributions include the design artifacts themselves, new foundations (constructs, models, methods, and instantiations), and new [evaluation] methodologies.
5. Research Rigour – Research methods must be rigorously applied.
6. Design as a Search Process – Research must be conducted with knowledge of other, competing approaches and should approach the process as a cyclical problem solving process, in which solutions are tested against each other and against their efficacy for solving the full problem.
7. Communication of the Research – Presentation of results needs to address both the rigour requirements of the academic audience and the relevance requirements of the professional (e.g. managerial) audience.

Interestingly, Hevner et al. [6] do not mention new or revised theory as a design artifact or research contribution (even though they reference Markus et al. [10]); it does not have a place in their guidelines.

Hevner et al. [6] emphasise Guideline 3, noting that "evaluation is a crucial component of the evaluation process" (p. 85). They further note that the evaluation method must be matched to the artifact and any evaluation metrics.

An important point about the Hevner et al. [6] guidelines is that the authors stated "Following Klein and Myers [8] we advise against mandatory or rote use of the guidelines." (p. 82).

2.3 Post Hevner et al. (2004)

Developed and written practically simultaneously with Hevner et at [6], the Design Research web pages [14], a portion of the AISWorld website, review much of the pre Hevner et al. [6] research and provide their own perspective on DSR. The website implicitly recognises a role for theory and theorising in DSR, identifying "better theory" as an important higher level of abstraction. However, it does not pick up the idea of an ISDT [18, 10].

Venable [15] made a case for the need to develop ISDTs, which he called "Utility Theories". He asserted that precisely formulated theories are a key vehicle for communication between scholars. He also asserted that a viable design theory does not really need kernel theories, testable hypotheses, of a design method [18]. Rather, he asserted that the essential part of an ISDT is simply a theory that some meta-design has utility for addressing some meta-requirement(s).

Expanding on Nunamaker et al. [11] and Venable and Travis [17], Venable [16] further developed the concepts of Evaluation in DSR, identifying two main classes of evaluation: Artificial and Naturalistic. He identifies naturalistic evaluation as being very important – "the proof of the pudding".

Gregor [4] developed a taxonomy of five theory types in IS. Type V, theory for design and action, is roughly synonymous with ISDT as in Walls et al. [18].

Gregor and Jones [5] expanded on the idea of theory for design and action, extending and refining the view of an ISDT in Walls et al. [18]. Their proposal for the anatomy of a design theory includes six essential (core) components, including purpose and scope, constructs, principles of form and function, artefact mutability, testable

propositions, and justificatory knowledge. It also includes two optional components, including principles of implementation and an expository. This structure is largely compatible with that of Walls et al. [18].

Baskerville et al. [2, 3] also further develop the idea of evaluation, noting difficulties with snapshot approaches to evaluation and simplistic understandings of organizational situations and events during evaluation. As a solution, they propose a "soft" design science which applies more intensive interpretive methods to evaluation.

In addition to work exploring DSR activities, deliverables, and guidelines, some recent work has specifically looked at the application of the guidelines from Hevner et al. [6].

Arnott and Pervan [1] examined how well Decisions Support Systems (DSS) publications prior to 2004 that used DSR approach fulfilled the Hevner et al. [6] guidelines. They assessed that evaluation was the biggest weakness, asserting that "Some form of convincing evaluation should [be] mandatory for design-science research." (n.p.) and noted that qualitative methods should be considered and used more often in evaluation. They also found that, as a practical matter, it was very difficult to assess whether guideline 6 (Design as a Search Process) had been addressed, unless the paper made specific efforts to explicate how they had done so. Finally, Arnott and Pervan [1] also found that the "level and quantity of theorizing" in DSS DSR papers "needs significant improvement" (n.p.). They suggested that guideline 4 (Research Contributions) could be broadened to include explicit contributions to theory.

Indulska and Recker [7] examined the extent to which DSR papers in the major IS conference literature since 2004 addressed the Hevner et al. [6] guidelines. They found that 36.8% of the DSR papers analysed merely stated that they followed DSR guidelines, 22.8% focused on one guideline, 7.0% focused on some, but not all guidelines, and 19.3% elaborated on the research's implementation of all guidelines. Remarkably, only 14.0% of the papers did *not* mention or explicitly demonstrate the use of the guidelines, demonstrating a strong expectation that they be followed.

2.4 Literature Summary

The views espoused in Hevner et al. [6] have come to dominate the criteria, standards, guidelines, and expectations for how DSR should be conducted and written about. There are, however, other voices and views, some of which are reviewed above. In particular, the issues of the amount and means for evaluation and the need for theory as an output are seen by some to be key open issues.

3 Research Methodology

To address the issues raised above and answer the research questions given in the introduction, a survey of DSR participants was designed and distributed and the responses received and analysed.

The survey respondents were selected to represent three classes of participant roles, (1) the gatekeepers to high quality journals, (2) the gatekeepers to DSR conferences, and (3) DSR authors. For (1), the editors-in-chief, senior editors, and associate editors of the IS scholars' basket of eight journals (MIS Quarterly, IS Research, the European

Journal of IS, The Information Systems Journal, the Journal of the AIS, the Journal of MIS, the Journal of Strategic IS, and the Journal of IT) were chosen. For (2), the program chairs and program committee members of the DESRIST conferences (2006-2010) were chosen. For (3), the authors of papers published at DESRIST in 2006-2009 were chosen. While not perfectly representative of all IS DSR participants, the three groups seemed to be a suitable theoretical sampling method. In all, 338 journal editors, 10 DESRIST chairs, 79 DESRIST program committee members, and 242 DESRIST authors were in the sample. Due to overlaps in these categories, a total of 595 people were in the sample.

The survey itself comprised two sections. The first section collected some further demographic information, although much was already available as the survey was not done anonymously. Key questions included whether the respondent had DSR responsibility for a journal (not necessarily in the basket of eight) and whether any papers authored or co-authored by the respondent (not necessarily published at DESRIST) were either *about* DSR or *using* DSR.

The second section asked respondents to rate the importance of various guidelines or criteria on a 0-10 scale. The 0-10 scale was used because it would be clear that an interval scale was being used so that averages would be meaningful, a 0-10 scale is commonly used in day-to-day human activity, it allows fine levels of gradation if desired, and it is easy to remember. The specific instructions to the respondents were:

> "Please rate your perception of the relative importance of the following existing or potential areas of practice and standards for Design Science Research. Please rate them from 0 to 10, where 0 indicates the practice or standard area is of no importance and 10 indicates that it is mandatory that the practice or standard must be met to consider publication of the Design Science Research result."

The response items in section two were divided into four sub-sections. Section A was concerned with the Hevner et al. [6] guidelines, Section B with evaluation methods and aspects to be evaluated, Section C with IS Design Theories and Section D with miscellaneous items. The specific text of each item is shown with the research findings further below. All sections also included areas for open comments.

An email address list of all potential respondents was developed from the conference papers, DESRIST websites and calls for papers, journal websites, searches of the AIS faculty directory, and general searches on the web. The survey was emailed to all recipients in mid January 2010 and a follow up was made about a week later. Emails that bounced were further researched and the survey re-sent, in some cases successfully. A second follow-up was made to non-respondents at the beginning of March 2010, which yielded about 50% more respondents.

In addition to completed surveys, the author received replies indicating that the respondent had insufficient knowledge of DSR to meaningfully answer the survey or did not review or handle DSR papers. In response to that, survey non-respondents were asked on follow-up to please notify the author if they felt that they were unqualified to answer the survey and/or did not have DSR journal responsibilities – without answering the remainder of the survey. Other respondents gave other reasons for not answering the survey, as described in the next section.

4 Survey Results

4.1 Demographic Results

Of the original 595 intended survey recipients, 34 recipients could not be contacted or were ineligible (no longer journal editors or the author of this paper), reducing the sample population to 561.

Of the 561 eligible survey recipients, 234 or 41.71% responded. However, of those, 125 (53.42% of respondents) indicated that they did not feel qualified to answer the survey. Five (2.14%) indicated that they had philosophical differences with the survey or DSR that precluded them answering the survey. Eleven (4.70%) indicated that they didn't have time or didn't do surveys. In total, 141 (60.26%) of the respondents did not answer the survey, which leaves 420 (561-141) surveyed for response rate purposes.

Of the 93 respondents who answered the survey (a response rate of 22.14% of the 420), 8 (1.90% of 420) did not answer all the quantitative questions and thus should not be included in the calculations, as all item ratings are relative to all other ratings.

Concerning the overall valid response rate, in all, 85 valid survey responses were received, yielding a response rate of 20.24% of the 420 eligible survey recipients (who had *not* indicated that they felt insufficiently knowledgeable to answer the survey, had incompatible philosophical positions or didn't have enough time to answer the survey). The author deems this response rate to be sufficient.

Of the 85 valid survey responses, 40.00% had a senior scholars' basket of 8 journal editor role, 31.76% served as a DESRIST PC chair or member, and 61.18% were DESRIST authors. Based on the survey responses, 37.65% indicated that they had DSR responsibility at a journal (not necessarily for a basket of eight journal), 54.12% had authored or co-authored a paper *about* DSR, and 74.12% had authored or co-authored a paper *using* DSR (not necessarily at DESRIST). Many respondents also supplied answers to the open comment questions, totaling about 15 pages of text.

4.2 Detailed Results

Tables 1 through 4 below show the detailed results of the ratings of the quantitative survey items. Higher ratings indicate higher importance, with 10 being mandatory or essential and 0 being completely unimportant or irrelevant.

As can be seen in Table 1 (next page), fulfilling *all* of the Hevner et al. [6] criteria was rated on average as only of medium importance, while specific guidelines were rated higher. Guidelines 1, 2, 3, and 4 were rated as very important, with guidelines 5, 7, and especially 6 being rated less so. Of the four individual types of artifacts, instantiations were rated as most important and models rated as least important. The standard deviation and minimum and maximum ratings also provide important information. All areas relating to the guidelines were rated by at least one respondent at 10, being essential or mandatory for publication, while guideline 2 had the highest minimum (five). Items with minima of 0 and maxima of 10 indicate a very high level of disagreement among at least some respondents. Standard deviations are fairly high for all guideline ratings, also indicating high levels of disagreement.

The open ended comments also provided much information and enlightenment. Several respondents commented that the guidelines are "too mechanistic", "a bit too

dogmatic", "a cookbook recipe" or "too simplistic to apply a guidelines checklist" and objected to their use as "mindless checklists". One even claimed that "Even Hevner et al. do not regard their guidelines as 'guidelines' anymore – rather as an evaluation instrument'. Despite the concerns raised, many respondents (often the same ones who raised concerns) noted that the guidelines were important and useful.

Many also noted that the importance of any one criterion or guideline is context dependent, e.g. depending on the kind of artifact developed, the stage of the research, or "the state of the art for the particular research area of the paper".

Table 1. Part A: Hevner et al. [6] Guidelines

Survey Item	Mean	Std Dev	Max	Min
Addressing ***all*** of the guidelines given in Hevner et al. [6] (described below)	5.36	3.00	10	0
Presenting an identifiable and viable design artifact (concept, model, method, or instantiation) as in March and Smith [9] (Guideline 1: Hevner et al., 2004)	8.39	1.78	10	3
Presenting one or more clearly defined new concepts [9]	6.52	2.47	10	0
Presenting one or more clearly explained new models [9]	6.16	2.41	10	0
Presenting one or more clearly explained new methods for building the artifact [9]	6.29	2.44	10	0
Presenting one or more example instantiations of the artifact [9]	7.30	2.23	10	0
Addressing a relevant and important problem (Guideline 2: Hevner et al., 2004)	9.05	1.21	10	5
Evaluating the utility, quality, and efficacy of the designed artifact (Guideline 3: Hevner et al., 2004)	8.31	1.61	10	2
Clearly identifying the novelty, generality, and significance of the contribution (Guideline 4: Hevner et al., 2004)	8.45	1.74	10	2
Rigourous application of the research methods (Guideline 5: Hevner et al., 2004)	7.33	1.90	10	0
Developing the design using a cyclical, problem solving search process (Guideline 6: Hevner et al., 2004)	6.09	2.46	10	0
Presenting the research to address both rigour for the academic audience and relevance for the professional audience (Guideline 7: Hevner et al., 2004)	7.20	2.07	10	1

Others commented on the unfulfillability of the criteria, especially in a single research paper or thesis. As one respondent stated,

> "The general problem I experienced when writing a DSR-paper is that it is practically impossible to address all the guidelines within one paper. If the artefact addresses a relevant problem for practice then normally the build/evaluate cycle is rather complex. But papers are limited to a certain amount of pages where one typically only can explain either the construction or the evaluation of the artefact in detail. However, in order to get accepted both parts have to be described. The result is that both construction and evaluation is described superficially. I think journals willing to publish DSR papers should know about this issue and differentiate between 'construction papers' and 'evaluation papers'."

A very few respondents objected to the "exclusive focus on technical artifacts" and "exclusion of context", reflecting a "narrow economic rationalist view of organizations" and "a narrow, functionalist nature of research".

With respect to the different types of artifacts, one respondent usefully noted that "I don't think that any artifact can be introduced (described) without constructs", thereby giving a clear rationale for why explication of constructs is (or should be) mandatory for reporting DSR.

Table 2. Part B: DSR Evaluation

Survey Item	Mean	Std Dev	Max	Min
Conducting some sort of evaluation of the designed artifact(s), whether artificial (not real world) or naturalistic (in a real setting) [15]	8.80	1.40	10	4
Conducting an Artificial evaluation [15] of the designed artifact, using such methods as a criteria-based evaluation, mathematical proof, computer simulation, role-playing simulation, or lab experiment	6.11	2.35	10	0
Conducting a Naturalistic evaluation [15], i.e., in the real world, with real users using a real instantiation of the design artifact to do real tasks, using such methods as a case study, field experiment, survey of users or other stakeholders, phenomenological or ethnographic study, or action research	7.18	2.16	10	0
Evaluating the *utility* of the designed artifact for *solving the problem to be addressed*	8.35	1.59	10	4
Evaluating the *efficiency* of the design artifact	6.35	1.88	10	0
Evaluating the *efficacy* of the designed artifact in a realistic setting	7.11	2.02	10	0
Quantitatively measuring the utility, efficiency, or efficacy of the designed artifact	5.74	2.42	10	0
Evaluating the designed artifact in comparison to other extant solutions to the problem	7.37	2.18	10	0
Evaluating the designed artifact for side effects (undesirable or desirable)	6.21	2.19	10	0

As shown in Table 2 above, at least some form of evaluation is rated as very important, consistent with the high rating of Hevner et al. [6] guideline 2 in Table 1. Of the two main kinds of evaluation, naturalistic evaluation is rated higher on average than artificial evaluation. Of the areas to be evaluated, utility for solving the problem to be addressed (i.e. meeting the meta-requirements in ISDT terms) is rated as more important than efficiency or efficacy. Of the other aspects of evaluation, rating in comparison to other extant solutions is rated as important, ahead of evaluation for side effects. Quantitative measurement during evaluation was rated the least important, but still of medium importance.

Many respondents commented on the essential, necessary nature of evaluation in DSR. Others were of the opinion that an implementation would need to exist before evaluation, so conceptual work might not be evaluated.

Some respondents noted that naturalistic evaluation is important, but also noted the difficulty in doing so ("resource intensive", "challenging") that the need for it depends on the nature of the designed artifact, and that it should not be mandatory.

One respondent noted that quantified measurements in evaluation increased the chance of publication, but reports of satisfaction from users of the instantiated artifact, especially from opinion leaders such as managers, were also persuasive to reviewers.

Overall, the respondents' comments reflected that the form of evaluation and what was to be evaluated could not be specified acontextually and should instead be selected carefully in line with resources, the stage of the research project or program, the kind of artifact, and the state of the art of the research area.

Table 3 (next page) shows the quantitative results of part C of the survey concerning the relative importance of IS Design Theories (ISDTs). Overall, the development and inclusion of all or part of ISDTs in DSR publications was rated on average as less important than fulfilling the individual guidelines in Hevner et al .[6], but not less important than meeting *all* of the Hevner et al. [6] guidelines and also less important than proper evaluation. Importantly, this area showed the highest variability in ratings, with maxima of 10 and minima of 0 for all items. This clearly reflects the divergence of opinion in the field concerning the mandatory view or irrelevant (nonsensical) nature of ISDTs.

The specification of meta-requirements [18] or purpose and goals [5] was rated the most important, followed closely by the specification of a meta-design [18] or constructs and principles of form and function [5]. The identification and relationship of kernel theories to the meta-design [18] and principles of implementation in specific contexts [5] were rated slightly less on average. Testable hypotheses were rated as the least important, but still of medium importance.

The open-ended comments raised several issues. Some respondents commented that the development and use of "design principles" was more relevant to them than "design theory". Related to this, others commented on the development of mid-range theories as being appropriate rather than full-blown design theories. Another respondent had a staged view of theory development, commenting that "a full and complete one may not be necessary, but a potential one that leads to a complete one should be important."

The difficulty in creating design theory also drew comments. One respondent wrote "Design Theory as of Walls et al. / Gregor&Jones is regarded as creating too much overhead - reduction to essential design theory seems to be useful". Another commented that there is a need for "a simple language for communicating theory".

Most importantly, other respondents indicated they had deep problems with design theory, in line with the division between March and Smith [9] and Walls et al. [18]/Gregor and Jones [5].

Table 3. Part C: IS Design Theories

Survey Item	Mean	Std Dev	Max	Min
Specifying a full and complete design theory, e.g. as in Walls et al. [18] or Gregor and Jones [5]	5.72	2.68	10	0
Specifying the meta-requirements for the generalised problem to be solved [18] or purpose and scope [5]	6.80	2.41	10	0
Specifying a meta-design (generalised design for meeting the meta-requirements, [18]) or principles of form and function for the design artifact product [5]	6.49	2.43	10	0
Specifying a design Method for instantiating the meta-design [18] or principles of form and function for the design artifact process [5]	5.99	2.28	10	0
Specifying kernel theory(ies) [18] or justificatory knowledge [5] relevant to how the meta-design meets the meta-requirements (Walls et al., 1992)	6.21	2.64	10	0
Specifying kernel theory(ies) [18] or justificatory knowledge [5] relevant to the design method (Walls et al., 1992)	5.95	2.41	10	0
Specifying testable hypotheses [18] or testable propositions [5] about how well the meta-design meets the meta-requirements (Walls et al., 1992)	5.79	2.72	10	0
Specifying testable hypotheses [18] or testable propositions [5] about how well the design method results in an artefact consistent with the meta-design [18]	5.73	2.64	10	0
Specifying constructs as representations of the entities of interest in the theory [5]	6.06	2.79	10	0
Specifying principle(s) of implementation in specific contexts [5]	6.09	2.46	10	0
Specifying an expository instantiation [5]	5.91	2.63	10	0

The findings of part D (Table 4 below) concern other miscellaneous guidelines. Relevance and significance of the problem and depth of analysis and clarity of understanding of the problem were rated as very important. Depth of analysis and clarity of problem understanding have not been addressed much in the literature, with the exception of Venable [15, 16]. Having a clear understanding of why an artifact works or doesn't work is also highly rated. Profoundness of insight and novelty are rated as important, but not as high. The size and complexity of the artifact and the effort that went into its development are rated as being of lesser importance, in fact the least important of all the items in the survey. Development effort and elegance were the only items that were not given a rating of 10 by any of the 62 respondents.

Table 4. Part D: Other Potential Criteria/Standards

Survey Item	Mean	Std Dev	Max	Min
Relevance of the problem to industry/society clearly established	8.05	1.77	10	1
Significance of the problem to industry/society clearly established	7.87	1.80	10	1
Depth of analysis and clarity of understanding of the problem and its causes	7.92	1.42	10	4
Depth or profoundness of insight leading to the new design artifact	7.35	1.61	10	3
Novelty of the new design artifact	7.29	2.02	10	0
Size and complexity of the new design artifact	4.51	2.36	10	0
Amount of effort that went into the development of the new design artifact(s)	4.25	2.29	9	0
Elegance of the design of the new artifact(s)	5.22	2.34	9	0
Simplicity of the design of the new artifact(s)	5.62	2.22	10	0
Clear understanding of why the new artifact works (or doesn't work)	7.68	2.04	10	0

Several respondents identified additional criteria. One area relates to the stakeholders' perspective(s), including "how easy it is for a user to understand, if it is "packaged" in a user-friendly way", "How usable the artifact is in addition to being useful. Will it burden the potential users instead of easing their effort?", "sensitivity to cultural and social contexts", "Stakeholder interests and analysis", "Impact in a real job/task of the design of the new artifact".

In another area, a few respondents highlighted the need to relate design theories more strongly to behavioural theories (i.e., related to kernel theories).

A third additional very criterion suggested is "Diffusion potential / economic potential (e.g. would you be willing to adopt the artefact? Would you pay for it?)". Market adoption is of course clear evidence of utility!

Finally, two respondents made suggestions that papers can contribute to improving DSR itself, e.g. through "novelty of the approach to Design Science Research" or "contribution to theoretical understanding and practical relevance of design science".

There were also very useful open-ended comments made. Some respondents justified low ratings for size and complexity and development effort. One stated, "A design artifact does not have to be complicated to be valuable. In addition, sometimes an individual may have a stroke of creativity and may arrive at a clever solution to a problem easily. The researcher should not be penalized for their 'amount of effort' nor should we encourage researchers to embellish their process to suggest that their artifact is indeed appropriate."

Other respondents reiterated their concern that guidelines need to be applied carefully according to the research context.

Finally, several respondents reiterated their concern with the whole idea of guidelines, lamenting how they are (inevitably) used, by reviewers and authors alike.

5 Discussion and Conclusions

In considering the first research question, this study clearly shows extensive disagreement on what guideline areas should be used as criteria and standards for evaluation of DSR. Nonetheless, there is near consensus on a few areas, such as the need to address and help solve an important problem, to have a clear design artifact, and to have some form of evaluation. Other areas, particularly the development and use of ISDTs are very controversial.

In considering the second research question, it is also clear that there are many views competing with the received view that the Hevner et al. [6] guidelines can or should be used as an evaluation checklist, even for top level journal publications.

First and foremost, many respondents cautioned against the use of the Hevner et al. guidelines (or any guidelines for that matter) as a mandatory checklist for evaluating DSR projects and publications. Further evidence is found in the ratings, with the rating of the survey item that *all* of Hevner et al. [6] guidelines *should* be met receiving only lukewarm support along with weak support for a few of the individual guidelines. The alternative perspective emphasises assessing whether the required rigour of the evaluation and the need to develop IS design theory are relevant and appropriate to the stage and scope of the research; more rigourous evaluation and ISDT development and validation are seen as appropriate for more mature DSR artifacts and not required for early stages of research.

Secondly, the ability of people to apply a fairly complex set of criteria is in more than a little doubt. Here, suggestions to simplify and clarify may be useful. On the other hand, perhaps one needs to recognise that reviewing and evaluating research is difficult and requires careful scrutiny and application of considered judgment.

Finally, some of the surveyed items outside of the Hevner et al. [6] guidelines received fairly average high ratings and additional areas for evaluating DSR were suggested in the open ended comments; these results conflict with a rebellion against mandatory criteria. A potential resolution that I suggest here is to use a cumulative model that adds up the value of the DSR work's contribution to some (but not necessarily all) of the various criteria, rather than the subtractive model inherent in a checklist approach (where all criteria not met fully count against the research). In such a model, the accumulated worth of the research might have a lower required level for acceptance in less rigourous publication venues. Only in top level journals (if indeed even there) would one consider that truly rigourous evaluation or theory development could be required and then only at later stages of the research and maturity of the artifact(s) as above. Moreover, if a high level of rigour is demanded, sufficient space needs to be allowed for explication of the motivation, development, description, evaluation, and theory outcomes of the research being reported.

A few limitations of the research are worth mentioning. First, as a first offering of the survey and initial analysis, the validity of the survey items may be questioned. Therefore, the specific numbers reflected in the averages should be considered as indicative at a high level rather than being precise measurements. Second, as pointed

out by some respondents, the idea that some guidelines can be assessed independently from others is also questionable.

So, where to from here? Overall, it seems clear that the diversity of opinion on the topic of criteria, standards, guidelines, and expectations for DSR needs significantly more discussion and refinement before researchers and reviewers can be comfortable with what needs to be done to produce high quality or even acceptable DSR. While Hevner et al. [6] is useful, it is not the be all and end all to the development of DSR; further improvement to the criteria, standards, guidelines, and expectations for DSR seems necessary. The proposal herein for a cumulative or additive model for assessing DSR may be useful. What level of cumulative value is appropriate and the relative importance of the different criteria also need further discussion. Clearly, more work on ISDTs and their role and value in DSR publications is also needed. The additional, non-Hevner-et-al guidelines suggested by this research also need more consideration and discussion. This paper is only one input to that discussion, which, presumably, should never be completely finalized, only reach a point of acceptable equilibrium before the next issue is raised and further change considered.

References

1. Arnott, D., Pervan, G.: An Assessment of DSS Design Science using the Hevner, March, Park, and Ram Guidelines. In: Information Systems Foundations: Answering the Unanswered Questions about Design Research. Australian National University (2008)
2. Baskerville, R., Pries-Heje, J., Venable, J.: Soft Design Science Research: Extending the Boundaries of Evaluation in Design Science Research. In: Chatterjee, S., Rossi, M. (eds.) Proceedings of the 2nd International Conference on Design Science Research in Information Systems and Technology (2007)
3. Baskerville, R., Pries-Heje, J., Venable, J.: Soft Design Science Methodology. In: Purao, S., Lyytinen, K., Song, I.-Y. (eds.) Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology. ACM Digital Library (2009)
4. Gregor, S.: The Nature of Theory in Information Systems. MIS Quarterly 30, 611–642 (2006)
5. Gregor, S., Jones, D.: The Anatomy of a Design Theory. Journal of the Association for Information Systems 8, 312–335 (2007)
6. Hevner, A., March, S., Park, J., Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28, 75–105 (2004)
7. Indulska, M., Recker, J.: Design Science in IS Research: A Literature Analysis. In: Information Systems Foundations: Answering the Unanswered questions about Design Research. Australian National University (2008)
8. Klein, H.K., Myers, M.D.: A Set of Principles for Conducting and Evaluating Interpretive Field Studies in Information Systems. MIS Quarterly 23, 67–93 (1999)
9. March, S., Smith, G.: Design and Natural Science Research on Information Technology. Decision Support Systems 15, 251–266 (1995)
10. Markus, M.L., Majchrzak, A., Gasser, L.: A Design Theory for Systems that Support Emergent Knowledge Processes. MIS Quarterly 26, 179–212 (2002)
11. Nunamaker Jr., J.F., Chen, M., Purdin, T.D.M.: Systems Development in Information Systems Research. Journal of Management Information Systems 7, 89–106 (1991)

12. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 3, 45–77 (2008)
13. Rossi, M., Sein, M.: Design Research Workshop: A Proactive Research Approach. Presentation delivered at IRIS 26 (2003), http://tiesrv.hkkk.fi/iris26/presentation/workshop_designRes.pdf
14. Vaishnavi, V., Kuechler, B.: Design Research in Information Systems, AISWorldNet, http://www.isworld.org/Researchdesign/drisISworld.htm
15. Venable, J.R.: A Framework for Design Science Research Activities. In: Proceedings of the 2006 Information Resource Management Association Conference (2006a)
16. Venable, J.R.: The Role of Theory and Theorising in Design Science Research. In: Hevner, A., Chatterjee, S. (eds.) Proceedings of the First International Conference on Design Science Research in Information Systems and Technology (2006b)
17. Venable, J.R., Travis, J.: Using a Group Support System for the Distributed Application of Soft Systems Methodology. In: Hope, B., Yoong, P. (eds.) Proceedings of the 10th Australasian Conference on Information Systems, Wellington, New Zealand, pp. 1105–1117 (1999)
18. Walls, J., Widmeyer, G.R., El Sawy, O.A.: Building an Information System Design Theory for Vigilant EIS. Information Systems Research 3, 36–59 (1992)