# Workshops as a Research Method:
# Guidelines for Designing and Evaluating Artifacts Through Workshops

Katja Thoring
Anhalt University of Applied Sciences,
Delft University of Technology
katja@thoring.com

Roland M. Mueller
Berlin School of Economics and Law,
University of Twente
roland.mueller@hwr-berlin.de

Petra Badke-Schaub
Delft University of Technology
p.g.badke-schaub@tudelft.nl

## Abstract

*Workshops are often used in the information systems (IS) and design fields to evaluate artifacts or to co-create business innovations. However, the evaluation of workshops is often conducted in a rather unsystematic and heterogenous way. This paper introduces a set of guidelines for designing or evaluating artifacts through workshops. These guidelines include five evaluation principles and a framework that outlines appropriate evaluation methods for different research goals. The relevant constructs and principles were identified based on related literature. The derived evaluation matrix was then revised based on ratings of five experts who independently assigned appropriate research methods for different evaluation foci. The framework's applicability was evaluated by comparing it with ten papers from the IS and design fields. The proposed guidelines can support researchers with conducting workshop evaluations in a comparable and replicable way, which will help to improve research rigor in the future.*

## 1. Introduction

Research method guidelines and principles are prevalent within the information systems (IS) community to provide structure and facilitation for conducting research in a rigorous way. Hevner [27:87] emphasized the need to "provide clear and consistent definitions, ontologies, boundaries, guidelines, and deliverables for the design and execution of high quality design science research projects" for the IS research community. Holtkamp, Soliman, and Siponen [29:6281] define research method principles as "any principles that provide normative guidance on how good research is conducted or evaluated (or both)."

Within the IS literature there exist numerous sources that suggested research method guidelines or principles for various purposes and contexts. Hevner et al. [28] presented seven guidelines of how to conduct design science research (DSR). Peffers, Tuunanen, Rothenberger, and Chatterjee [44] proposed a six-step design science research methodology (DSRM). Gregor and Jones [25] suggested eight components of a design theory in IS, that is, how to design an IS artifact. Sein, Henfridsson. Purao, Rossi, and Lindgren [48] suggested a four-stage cycle of action design research (ADR) that consists of seven principles. Siau and Rossi [49] reviewed techniques for evaluating design modelling methods. Durcikova, Lee, and Brown [17] introduced statistical action research, along with five principles (called heuristics), which suggest to include statistical hypothesis testing and sample surveys to the action research (AR) planning and evaluating stages. Klein and Myers [33] introduced a set of principles for interpretative field research. Venkatesh, Brown and Bala [56] presented a set of guidelines for mixed method research in IS.

While diverse in aims and approaches, these papers all demonstrate the relevance and need of normative, prescriptive guidelines for the IS community. But also, in the design innovation field, normative guidelines can be found. For example, Goffin, Åhlström. Bianchi, and Richtnér [23] suggested a set of ten quality evaluation criteria for case study research.

Design science research (DSR) typically deals with the development and evaluation of an artifact. Artifacts in this sense are not limited to physical or digital products, but also include intangible models, service systems, and methods [28].

Often, those artifacts are developed or evaluated in workshop settings. Hence, we argue that workshops are a relevant phenomenon of inquiry in the IS field. However, until now, there exist no such normative guidelines for evaluating DSR workshops. As emphasized by [28], DSR artifacts need "rigorous methods in both the construction and evaluation of the design artifact". The range of approaches, participants, sizes, and goals is quite broad. Therefore, a set of standardized guidelines for conducting workshops as an evaluation method or to design artifacts in a co-

HĮCSS

creation approach would ensure research rigor and comparability between different workshop results, as well as replicability of approaches. Developing and applying such guidelines is the goal of this paper.

*RQ: What guidelines for developing and evaluating design science artifacts through workshops can be derived from literature and expert inquiries?*

We focus solely on workshops to develop and evaluate DSR artifacts. Other types of workshops, such as scientific conference symposia or educational training workshops, are beyond the scope of this paper.

In Section 2, we start by discussing relevant related work that guided our deductive guideline development. Section 3 outlines our methodology and approach for the guideline development and application. In Section 4, we present five evaluation principles for DSR workshops and a related matrix for identifying appropriate research methods. In Section 5, we evaluate the applicability of our suggested guidelines through comparisons with exemplary papers from the IS and design fields. In Section 6, we discuss our approach and conclude by outlining limitations and future work.

## 2. Theoretical Framework

March and Smith [39] differentiate between two design processes produced by design-science research in IS: build and evaluate. In a similar vein, a workshop can have the goal to either create new output (ideas, concepts, designs), or to evaluate specific aspects of interest, such as testing the usefulness of a designed product, process, or tool. In the latter case, the evaluation of the workshop is part of the core research question and the workshop is conducted with the sole purpose to evaluate something else. But also, the first case, such as design thinking or other practice workshops with the purpose to produce some kind of innovative idea, might need to be evaluated, for example, to analyze the workshop's outcome or to check whether the agenda and timing of the workshop were appropriate.

Venable et al. [55] introduced a framework for evaluation in design science (FEDS). It differentiates between two dimensions: the functional purpose (formative vs. summative) and the evaluation paradigm (artificial vs. naturalistic settings). Workshops could be used in all four quadrants, but we argue that workshops are most useful in formative naturalistic settings, because they typically test the risk for user acceptance in context rather than pure technical aspects.

Often, evaluation workshops are part of an action research [3, 32, 36, 42, 52, 58] or action design

research [48] project. However, the literature on both does not suggest concrete, prescriptive guidelines on how to evaluate those workshops.

Also, literature about workshops as a research method in the design and IS fields is scarce.

Tremblay et al. [54] presented guidelines for how to adapt focus groups for design research. They distinguish between exploratory and confirmatory focus groups. An exploratory focus group seeks to improve an artifact, while a confirmatory focus group would judge an artifact. This distinction is similar to formative and summative evaluation [55]. We consider focus groups as only one possible method in a workshop to design or evaluate an artifact.

Storvang, Mortensen, and Clarke [51] presented some ideas on how to plan, diagnose, facilitate, and analyze a workshop, however, without providing concrete guidelines or principles for the evaluation phase.

We acknowledge that workshops are diverse and multifarious in nature, and hence, cannot be easily generalized. Therefore, we decided not to include any guidelines for planning and conducting a workshop, but to focus on the task of designing and evaluating artifacts through a workshop, only.

Different workshop parameters can be also tested in an experimental fashion. However, workshops are to be distinguished from classical controlled laboratory experiments because they are more qualitative and the participants work collaboratively for a common workshop outcome (e.g. ideating a data strategy or designing a business process).

## 3. Methodology

Our research approach can be structured into three steps: (1) identifying relevant constructs of an evaluation workshop, (2) developing the guidelines, and (3) evaluating the applicability of the guidelines.

### 3.1. Methodology for identifying relevant constructs and principles

Following a deductive approach, we identified relevant constructs and principles for workshop evaluations, as presented in relevant literature.

First, we reviewed relevant papers that presented research method principles in other areas [25, 28], and adapted those suggested principles for the specifics of workshop evaluations. Also, the steps of the action research process [15, 32, 36] were consulted in order to formulate five principles for workshop evaluations.

Second, we reviewed the literature on workshops in general and regarding workshops in the educational sciences field [4, 13, 34, 43, 51]. Here, we found useful information that led to the identification of relevant

constructs for workshop evaluation goals, that is, different possible purposes and intended outcomes of a workshop.

Finally, we reviewed various sources about research methods in general and mixed methods and triangulation in particular [5, 12, 29, 31, 56], in order to identify relevant methods for workshop evaluations. The results of this step are described in Section 4.1.

## 3.2. Methodology for developing the workshop evaluation framework

After defining the relevant constructs for the workshop evaluation (evaluation goals, and evaluation methods), we consulted five experts to rate the appropriateness of specific research methods for the respective evaluation goal. The experts had substantial professional experience with planning, conducting, and evaluating workshops in different contexts. We selected the experts to cover a broad range of areas of expertise, including IT, management, and design. Among the five experts, two are co-authors of this paper, while the other three were included to provide an independent perspective. Table 1 outlines the five experts and their experience and areas of expertise.

**Table 1. Overview of consulted experts.**

| # | Experience (# Years) | Conducted Workshops | Area |
|---|---|---|---|
| 1 | 30+ | 150+ | Strategy and Management |
| 2 | 20+ | 70+ | Strategic Design |
| 3 | 5+ | 30+ | Digital Product Development, Data Strategy |
| 4 | 15+ | 50+ | Design Thinking |
| 5 | 15+ | 50+ | IT Strategy and Innovation |

The five experts were given an empty matrix including the previously identified constructs. One dimension of the matrix included the respective workshop goals, that is, which aspects should be evaluated. The other dimension contained a list of qualitative and quantitative research methods. The experts then rated the usefulness of each method for each workshop goal (good, medium, low). Subsequently, we compared the ratings of the five experts and calculated a Fleiss' Kappa [19] inter-rater agreement coefficient of 0.27, which is categorized as "fair" [35]. Additionally, we used the Spearman's statistics for calculating the pairwise correlation among raters. We calculated the mean of all correlations between each possible pair of raters and received a Spearman's correlation of $\rho = 0.46$.

These relatively low agreement coefficients demonstrate the core of the problem: many researchers seem to have different understandings of appropriate research methods for specific workshop goals and evaluation methods, which warrants our attempt to develop such guidelines.

To develop the final framework, as shown in Table 2, we differentiated between *strong* disagreements (two steps distance—one rating said "good", another said "low") and *minor* disagreements (one step distance—e.g. one rating said "good", another said "medium"). For the minor disagreements, we chose the rating that was suggested by the majority of researchers. For all remaining strong disagreements, the ratings were discussed among the five experts, until an agreement was found. The resulting workshop evaluation framework, that was agreed upon by the five experts, is described in Section 4.2 and shown in Table 2.

## 3.3. Methodology for applying the workshop evaluation guidelines

In order to evaluate the applicability of the developed guidelines, we compared them with selected papers from the IS and design fields that involved some sort of workshop evaluation.

To identify relevant papers, we conducted a literature search in the Scopus database with the keywords "workshop" AND ("evaluation" or "analysis"). The returned sources were limited to the AIS "basket of eight" journals [1], which returned four papers, (2) to the main AIS conferences (HICSS, AMCIS, ICIS, ECIS, and PACIS), which returned 52 papers, and (3) to the top journals of the design field [20], which returned 75 papers. The AIS conferences were included, because the "basket of 8" AIS journals did not yield a significant number of relevant papers.

The resulting 131 papers were screened based on their title and abstract in order to identify relevant papers (that is, papers that described an actual evaluation workshop) and to exclude redundant and unrelated papers (e.g. conference papers that were later adapted for a journal publication, or papers that reported on conference tutorials rather than applied research workshops). From the remaining papers, we selected 10 to include in our final analysis. These 10 papers were chosen to include a variety of (1) different sources (specifically, to include papers from different design journals, different AIS journals, and different AIS conferences), and (2) a variety of different workshop evaluation goals, in order to compare those with our suggested framework.

The selected papers were not meant to provide a representative sample from the included sources, but rather a snapshot of heterogenous studies, in order to illustrate the applicability of our suggested framework.

The results of this step are described in Section 5 and illustrated in Table 3.

# 4. Workshop evaluation guidelines

This section summarizes the developed workshop evaluation guidelines. Following Holtkamp et al. [29], we use the term "guidelines" for the overall set of different principles and tools for evaluating workshops. More specifically, we present several workshop evaluation "principles" (Section 4.1), workshop evaluation "goals" (Section 4.2), workshop evaluation "methods" (Section 4.3) and a workshop evaluation "framework" that juxtaposes appropriate methods for specific goals. All four aspects, that form our suggested workshop evaluation guidelines, are discussed in the following sections.

## 4.1. General principles

Based on the previously discussed related work (Section 3.1), we suggest the following five general principles for evaluations of DSR workshops.

1. **Focus Definition.** Define a concise research question and evaluation goals. What *exactly* should be achieved through the workshop? Refer to Figure 1 to identify areas of interest. Is your focus on evaluating something or on creating something?

2. **Role allocation.** Identify roles among involved stakeholders. Determine which roles should serve as informants for what information. Minimize the risk of researcher bias, for example by separating coaching from data collection.

3. **Triangulation.** Combine various research methods. Compare data from more than one appropriate data source, according to Table 2.

4. **Transparency**. Describe and publish your evaluation goals, evaluation methods, selection criteria, participants' details, workshop course, and workshop results, to allow replicability for other researchers.

5. **Reflection.** Formulate 3–5 main insights about the usefulness of your evaluation *procedure* (not the results). This will help to further improve the workshop evaluation guidelines in the future and provide helpful insights for other researchers.

## 4.2. Evaluation goal

Building on the work of Storvang et al. [51] we identified several relevant categories that inform a workshop setup, including the workshop's goal, procedure, and outcome, involved people and their actions, involved facilitation tools, and introduced artifacts. Figure 1 illustrates the different aspects of a workshop that could be evaluated. More specifically, we distinguish between seven relevant aspects. (A) The workshop goal needs to be defined clearly in the beginning. In DSR the evaluation goal is typically either a new artifact that is developed during the

workshop in a co-creation approach, which would be the workshop's outcome (G), or the evaluation of an existing artifact that is supposed to be tested or improved through the workshop. This artifact could be either a previously designed introduced artifact, such as a product, a software, a service, or a process (B). Also, facilitation artifacts, such as tools, templates, or the workshop space could be the focus of interest and be evaluated (C). Besides these main evaluation objectives, also secondary aspects could be investigated through the workshop. More specifically, the workshop procedure, such as agenda and timing (D) can be evaluated, and the people's roles (E), as well as their dialogs, opinions, thoughts, and interactions with the artifact (F) can be of relevance. The secondary aspects are marked in grey in Figure 1.

Each of these aspects requires different approaches and methods for evaluation. In the following section we present a framework of guidelines for each workshop goal, developed based on the five experts' ratings (as described in Section 3.2).
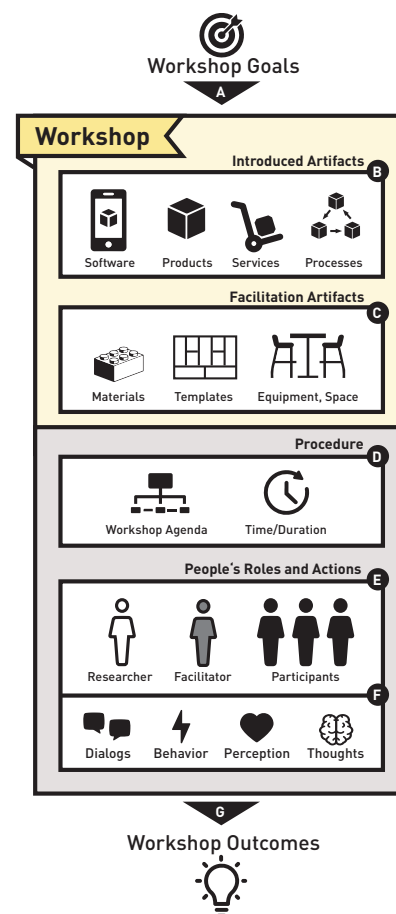


**Figure 1. Overview of possible workshop evaluation goals**

**Table 2. Goal–Method framework for DSR workshops, based on consensus of five experts.**

| | Observation & Notes | Photography | Video Recording | Audio Recording | Survey & Questionnaire | Interview | Group Discussion | Artifact Analysis |
|---|---|---|---|---|---|---|---|---|
| Introduced Product, Software, Service, or Process (Artifact) | ● | ◑ | ● | ○ | ● | ● | ● | ◑ |
| People's Behavior (during workshop) | ● | ◑ | ● | ○ | ○ | ◑ | ◑ | ○ |
| People's Dialogues (during workshop) | ◑ | ○ | ● | ● | ○ | ○ | ◑ | ○ |
| People's Interactions (during workshop) | ● | ◑ | ● | ○ | ○ | ◑ | ◑ | ○ |
| People's Perception and Feelings | ○ | ○ | ◑ | ◑ | ● | ● | ● | ○ |
| People's Opinion and Ideas | ○ | ○ | ◑ | ◑ | ● | ● | ● | ◑ |
| Workshop Agenda and Timing | ● | ○ | ◑ | ○ | ◑ | ● | ● | ◑ |
| Workshop Equipment, Furniture, Space | ● | ◑ | ● | ○ | ◑ | ◑ | ● | ◑ |
| Workshop Materials, Tools, Templates (e.g. Canvas, | ● | ● | ● | ○ | ● | ● | ● | ● |
| Workshop Outcome Quality (e.g. Ideas, Strategy) | ◑ | ○ | ◑ | ○ | ● | ● | ● | ● |

○ = low suitability    ◑ = medium suitability    ● = good suitability

## 4.3. Evaluation methods

We distinguish different evaluation methods as applicable to design or evaluate artifacts through workshops. Each of them has their merits and perils. Table 2 presents the developed goal–method framework that can be used to guide workshop evaluations.

Observations are highly useful for analyzing how people behave and how they interact with each other or with provided tools and artifacts. They allow to investigate people in an unobtrusive way in real-time. In a workshop setting, the researcher can become part of the situation and hence gain first-hand insights, for example by acting as a coach. This would resemble participant observation, known from ethnography, where researchers would live among the group to be studied for a certain period of time [18:45]. However, it needs to be taken into account that this situation bears a high risk of bias, because the coach can easily loose his/her independent role. When coaching a team, the researcher can influence the team's behavior, implant his/her own ideas, or direct the group towards his/her intended research outcome—be it consciously or unconsciously. Observation data is typically captured through researchers' field notes. This technique is also prone to researcher bias, because some sort of interpretation might already take place while notetaking. Moreover, researchers might be distracted while taking notes and hence miss other relevant activities happening during notetaking. Observations can also include some sorts of measurements (e.g. time spent on specific tasks).

Photography is another form of research method to capture observation data visually [11]. Pictures taken during the workshop can capture various aspects of the workshop, such as group activities, interactions with artifacts, or resulting workshop outcomes, such as filled whiteboards, created models, or written Post-it notes. The pictures can be later analyzed by the researchers. Hence, photography can be used to facilitate artifact analysis. However, pictures can provide only a snapshot of the real scene and never give an authentic picture of the whole workshop. When analyzed after the workshop, important aspects that are not visible in the pictures, might get lost. Therefore, photography should only be used as an addition to other research methods.

Video analysis is similar to observations, but allows to observe people's behavior and interactions also post-hoc, more than once, and by independent people who had not participated in the workshop [9]. Consequently, video observations can yield more details and secure higher accuracy, because the recorded data can be rewound indefinitely. Moreover, it can help to avoid possible researcher bias, because it can be analyzed by independent researchers. Ideally, the workshop would be recorded from various perspectives, including details of particular groups, as well as the entire scene from two different angles [9]. However, video observations are more obtrusive than personal observations and, hence, people might feel disturbed and behave differently [38]. Also, certain ambient characteristics of the workshop (such as atmosphere, smells, personal chats, or other contextual details) might get lost. Moreover, researchers should consider the high effort that is required for analyzing video data.

Audio data is particularly helpful when one is interested in people's dialogs during the workshop, or to analyze expressed feelings, thoughts, and ideas, for

example during presentation rounds. Audio recording is less obtrusive than video, but needs also specific consent upfront. However, important context information will get lost, when relying on audio, only, such as gestures, facial expression, or reactions of the audience. Audio recording is standard procedure to augment interviews, but in that case, it would not be considered an individual research method on its own, but would become part of the interview.

Interviews with workshop participants or other stakeholders (coaches, researchers) are particularly useful to investigate people's experiences [8:49], as well as their thoughts, ideas, opinions, and perceptions about introduced or facilitating artifacts, tools, equipment, products, processes, services, or software. However, interviews that are usually conducted after the workshop, will only rely on people's memories, which might not be very accurate, depending on the elapsed time. Moreover, the effort for analyzing interviews is relatively high.

A group discussion with participants (similar to a focus group [5:111, 54]), which represents some kind of workshop in itself, can also be used to evaluate artifacts. The advantage, compared to individual interviews, is the more complex level of discussions. Different opinions can be discussed immediately. On the downside, focus groups can result in averaging of opinions, because people tend to get influenced by their peers. Extreme opinions will most likely be diminished. It is important to differentiate between focus groups as a research workshop, and focus groups as a workshop evaluation method. When using a group discussion to evaluate a (previously conducted) workshop, the discussion needs to be moderated, and verbalized opinions need to be captured appropriately, for example by taking notes or through audio or video recording. When using focus group workshops as a research method, one should refer to the framework presented in Table 2 to identify appropriate evaluation methods according to the respective research questions.

Surveys and questionnaires can be included into the workshop evaluation process. They can be conducted shortly after the workshop by email or phone [16] or even be integrated into workshop tools (such as a canvas). Triangulation of such quantitative methods with the usually more qualitative approach of the workshop adds the opportunity to gain measurable feedback in a standardized and, hence, comparable form. Questionnaires allow to identify different opinions and perceptions between different user roles, and to allocate specific feedback to individual teams and their workshop results. Other than, for example, focus groups, questionnaires give every participant the possibility to express their opinion. Moreover, they allow to inquire feedback about the possible long-term impact of the workshop after some time has passed (which would also be possible through interviews and focus groups, but require less time and effort). On the downside, surveys and questionnaires allow to only inquire individual opinions and perceptions and are usually not that helpful to gain more complex feedback. Open-ended questions provide only limited possibilities for such detailed feedback. Briggs et al. [7] developed the Satisfaction Attainment Theory (SAT) and a corresponding survey instrument to measure outcome and process satisfaction of meetings.

Artifacts that are generated or modified during the workshop by the participants, provide a valuable resource for later analysis [45:91, 46]. In the context of a workshop, such artifacts can include, for example, filled canvases, models created by the participants, and labelled Post-it notes. Also, the workshop space itself, and how people use it, re-decorate it, or leave it after the workshop, can provide insights about specific behaviors or aspects of the environment. This kind of artifact analysis is a research method adapted from archaeology and material culture studies [10], in which archaeologists inferred insights about ancient cultures' lives from found objects and fragments.

## 4.4. Framework of possible workshop evaluation approaches

Table 2 outlines the possible choices of methods for the different research goals, based on the five experts' ratings. According to the question of interest that should be evaluated through the workshop, the appropriate research methods can be identified and included for data triangulation.

Combinations of several evaluation methods for the same evaluation goal can increase research validity through triangulation (e.g. for analyzing people's interactions, a triangulation of observations and video analysis is suggested). But also, combinations within the evaluation goals can be necessary. For example, if the focus of interest is on how people interact with a given tool, the suggested evaluation methods for both—tools, and people's interactions—should be considered.

We would like to emphasize that the suggested framework is not meant to be strictly applied but rather seen as a help to identify possible options. The final choice of methods is dependent on various factors and should be made by the researchers with careful consideration. However, we suggest to explain the choices for or against a specific method (as suggested by our general workshop evaluation principle 4, presented in Section 4.1).

**Table 3. Selected papers for testing the applicability of the framework.**

| Author | Title | Evaluation Goal | Evaluation Methods | Suggested by Framework |
|---|---|---|---|---|
| Irani et al. [30] | Electronic transformation of government in the U.K. | (1) People's Opinions and Ideas | (1) Audio Recording, (2) Observational Notes, (3) Artifact Analysis | (1) Survey & Questionnaire, (2) Interview, (3) Group Discussion |
| Sankar et al. [47] | An evaluation of a workshop with a focus on fostering teaching excellence through research | (1) People's Perceptions, (2) Workshop Materials, Tools, Templates | (1) Survey & Questionnaire | (1) Survey & Questionnaire, (2) Interview, (3) Group Discussion, (4) Observation & Notes, (5) Photography, (6) Video Recording, (7) Artifact Analysis |
| Giessmann [21] | Do end-users accept end-user development? | (1) People's Perceptions, Attitudes, (2) People's Opinion, (3) Introduced artifact | (1) Survey & Questionnaire | (1) Survey & Questionnaire, (2) Interview, (3) Group Discussion |
| Nagle and Sammon [41] | The data value map: A framework for developing shared understanding on data initiatives | (1) People's Perceptions (2) People's Opinions and Ideas, (3) People's Interactions, (4) Workshop Materials, Tools, Templates | (1) Observations, (2) Survey & Questionnaire, (3) Artifact Analysis, (4) Interviews | (1) Survey & Questionnaire, (2) Interview, (3) Group Discussion, (4) Observation & Notes, (5) Video Recording, (6) Photography, (7) Artifact Analysis |
| Morana et al. [40] | Designing a process guidance system to support user's business process compliance | (1) Introduced Artifact, (2) People's Opinion | (1) Group Discussion, (2) Audio Recording (3) Observational Notes (4) Artifact Analysis | (1) Observation & Notes, (2) Video Recording, (3) Survey & Questionnaire, (4) Interview, (5) Group Discussion |
| Gilsing et al. [22] | A reference model for the design of service-dominant business models in the smart mobility domain | (1) Introduced Artifact | (1) Audio Recording, (2) Observation & Notetaking, (3) Group Discussion (4) Artifact Analysis | (1) Observation & Notes, (2) Video Recording, (3) Survey & Questionnaire, (4) Interview, (5) Group Discussion |
| Austin et al. [2] | Mapping the conceptual design activity of interdisciplinary teams | (1) Introduced Artifact, (2) People's Interactions, (4) Workshop Materials, Tools, Templates | (1) Survey & Questionnaire (2) Observation & Notes (3) Audio Recording (4) Survey & Questionnaire, (5) Artifact Analysis | (1) Observation & Notes, (2) Video Recording, (3) Survey & Questionnaire, (4) Interview, (5) Group Discussion, (6) Photography, (7) Artifact Analysis |
| Dekoninck et al. [14] | New tools for the early stages of eco-innovation: an evaluation of simplified TRIZ tools | (1) Workshop Materials, tools, Templates (2) People's Opinions and Ideas | (1) Observations, (2) Artifact Analysis, (3) Photography (4) Group Discussion (5) Survey & Questionnaire | (1) Observation & Notes, (2) Photography, (3) Video Recording, (4) Survey & Questionnaire, (5) Interview, (6) Group Discussion, (7) Artifact Analysis, (8) Interview |
| Sohn and Nam [50] | Understanding the attributes of product intervention for the promotion of pro-environmental behavior | (1) Introduced artifact, (2) Workshop Materials, Tools, Templates | (1) Interview, (2) Group Discussion, (3) Audio Recording | (1) Observation & Notes, (2) Video Recording, (3) Survey & Questionnaire, (4) Interview, (5) Group Discussion, (6) Photography, (7) Artifact Analysis |
| Villalba [57] | Testing literature-based health experience insight cards in a healthcare service co-design workshop | (1) Workshop Materials, Tools, Templates, (2) People's Interactions, (3) People's Dialogues | (1) Audio Recording, (2) Video Recording, (3) Artifact Analysis, (4) Photography | (1) Observation & Notes, (2) Photography, (3) Video Recording, (4) Survey & Questionnaire, (5) Interview, (6) Group Discussion, (7) Artifact Analysis, (8) Audio Recording |

# 5. Application of the workshop evaluation guidelines

In order to evaluate the applicability of the suggested guidelines, we compare them with ten exemplary papers from the design discipline and the IS field. Table 3 outlines the respective workshop evaluation goals, and the employed evaluation methods (if any) for the ten analyzed papers.

Our analysis shows that many of the papers made only limited use of possible evaluation methods. Some papers relied on only one method and did not employ method triangulation [21, 47]. Also, the potentials of video analysis were seemingly not well considered. Only one paper [57] used video for a workshop evaluation, although this technique was rated as one of the most useful evaluation methods by our experts (see Table 2). Other papers employed different methods than those we have suggested in our framework [30].

To conclude, we argue that our presented guidelines would help to get a broader perspective of possible evaluation methods and hence, possibly improve the research rigor of evaluation workshops.

# 6. Discussion

## 6.1. Scope

A workshop is usually integrated into a broader research context. Often the workshops in the analyzed papers were a part of a design science, action research,

or action design research context. This paper is solely focusing on the evaluation of a research workshop. Other relevant aspects within the broader research project are not within the scope of this paper. Among these aspects are, for example, the design of products and workshop facilitation tools, briefing of coaches, the design of workshop input (e.g. presentation slides), pilot studies, etc.

Also, the analysis of the data collected from the workshop is beyond the scope of this paper. We argue that such data analysis procedures are not peculiar to workshops but would be the same as for any other data analysis task. The interested reader can refer, for example, to general literature about the analysis of qualitative research data [6]. Moreover, there exist various sources on specific data analysis topics, such as "Qualitative Data Analysis" (QDA) software, that can aid with the coding and clustering of audio, video, interview, and photographic data [37]. Protocol analysis refers to the systematic, step-by-step recording of activities and verbal outputs for individual and group processes [26]. Similarly, linkography [24] suggests a systematic graphical analysis process for team processes. And finally, new technologies, crowdsourcing, or data mining can be employed for analyzing large amounts of data [53].

## 6.2. Implications

Our literature search (Section 3.2.) yielded 59 papers from AIS conferences that present some sort of evaluation workshop. This fact indicates that workshops as a research method are of general interest to the IS field. However, only four papers were found within the "basket of 8" AIS journals. One could argue that studies involving workshop evaluations do not (yet) meet the quality standards and rigor that are required by the top IS journals. If this is the case, the guidelines presented in this paper might help to increase research rigor in the future.

This paper has implications for research by suggesting (1) how to evaluate workshops in a scientific rigorous way and (2) how to present the research results in a transparent, accessible, and replicable way. The presented framework juxtaposes different possible workshop goals with the appropriate evaluation methods. This framework could serve as a decision matrix for authors, editors, and reviewers.

Our analysis of the IS papers showed that up to now, workshops are often evaluated in an ad-hoc manner and that the choices of evaluation methods are limited, compared to the design discipline.

Furthermore, the presented guidelines can also be of relevance for practice. Workshops are widely used in practice, for example in the design thinking field, either to brainstorm ideas or to test user acceptance of

developed prototypes. We argue that our guidelines can be of help for practitioners who want to conduct workshops to evaluate an artifact or to design new ones.

## 6.3. Limitations and future work

Recently, some criticism against research methods guidelines or principles has been raised, because such guidelines could be misused as some sort of checklist to reject paper submissions, in cases where the respective guidelines were not sufficiently met [29]. We agree, that the relevance of a paper's content should be more important than the strict following of research method guidelines, but still we argue that such guidelines can help to maintain a rigorous research process and to enhance comparability. Consequently, we follow Holtkamp et al. [29] by suggesting that our developed guidelines are supposed to help researchers with their workshop evaluations but should not be used as a mere checklist without reflecting upon their appropriateness. Although we argue that our guidelines will ensure a higher standard of workshop evaluations, we acknowledge that there might be reasons for not following them strictly, such as limitations of resources or the fact that only the researchers would be able to also act as facilitators of the workshop for reasons of unique expertise. In such cases, we suggest to explain those limitations accordingly.

Moreover, Holtkamp et al. [29] criticized that many of the existing research method guidelines in the IS field were not tested themselves. We considered this demand by integrating five experts and consulting relevant literature for developing the guidelines, as well as applying the guidelines on a total of ten selected studies from the IS and design fields. We would like to emphasize, though, that the introduced goal–method framework is considered a first step that was developed based on the consensus of five experts. Future work will include further assessments with more experts in order to validate the framework.

Another goal for future work includes the refinement of the framework regarding the intended purpose of the workshop. Some methods might be more appropriate to design an artifact, while others are better for either summative judging or formative improving an artifact [55]. This differentiation needs further investigation.

## 7. Conclusions

This paper presents a set of research method guidelines for designing or evaluating DSR artifacts through workshops. Our guidelines include a list of five evaluation principles and a framework of possible research methods for different evaluation goals.

We describe the development process of the suggested guidelines and investigate their applicability by means of ten papers from the IS and design fields.

We consider the guidelines presented in this paper as a first step to instigate discussion about possible evaluation techniques of research workshops. Future work will include further testing and possible modifications of the guidelines. This goal could be achieved, for example, by presenting the guidelines to other researchers and practitioners who are planning to conduct workshops, and to analyze whether the guidelines would change their planned approaches or impact the outcome of the workshop evaluations.

## 8. References

[1] Association for Information Systems, "Senior Scholars' Basket of Journals", https://aisnet.org/page/SeniorScholarBasket

[2] Austin, S., J. Steele, S. MacMillan, P. Kirby, and R. Spence, "Mapping the conceptual design activity of interdisciplinary teams", *Design Studies 22*(3), 2001, pp. 211–232.

[3] Avison, D.E., F. Lau, M.D. Myers, and P.A. Nielsen, "Action research", *Communications of the ACM 42*(1), 1999, pp. 94–97.

[4] Ayers, J.B., *Evaluating Workshops and Institutes*, Office of Educational Research and Improvement, Washington, 1989.

[5] Berg, B.L., H. Lune, and H. Lune, *Qualitative research methods for the social sciences*, Pearson Boston, MA, 2004.

[6] Boeije, H., *Analysis in Qualitative Research*, SAGE, Los Angeles, 2009.

[7] Briggs, R.O., G.-J. de Vreede, and B.A. Reinig, "A theory and measurement of meeting satisfaction", *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, IEEE (2003).

[8] Brinkmann, S., *Qualitative interviewing*, Oxford University Press, Oxford, 2013.

[9] Candy, L., Z. Bilda, M.L. Maher, and J.S. Gero, "Evaluating software support for video data capture and analysis in collaborative design studies", *QualIT conference, University of Sydney, Australia*, (2004).

[10] Cochran, M.D., and M.C. Beaudry, "Material culture studies and historical archaeology", In M.D. Cochran and B. Beaudry, eds., *The Cambridge Companion to HistoricalArchaeology*. Cambridge University Press, Cambridge, 2006, 191–204.

[11] Collier, J., and M. Collier, *Visual Anthropology: Photography as a Research Method*, UNM Press, Albuquerque, 1986.

[12] Creswell, J.W., *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, SAGE, Thousand Oaks, 2013.

[13] Davis, L.N., and E. McCallon, *Planning, Conducting, Evaluating Workshops. A Practitioner's Guide to Adult Education*, Learning Concepts, Austin, 1975.

[14] Dekoninck, E., D. Harrison, and N.A. Stanton, "New tools for the early stages of eco-innovation: an evaluation of simplified TRIZ tools", *Journal of Design Research 6*(1–2), 2007, pp. 122–141.

[15] DeLuca, D., M.J. Gallivan, and N. Kock, "Furthering information systems action research: A post-positivist synthesis of four dialectics", *Journal of the Association of Information Systems 9*(2), 2008, pp. 48–72.

[16] Dillman, D.A., J.D. Smyth, and L.M. Christian, *Internet, mail, and mixed-mode surveys: the tailored design method*, John Wiley, Hoboken, NJ, 2014.

[17] Durcikova, A., A.S. Lee, and S.A. Brown, "Making rigorous research relevant: Innovating statistical action research", *MIS Quarterly 42*(1), 2018, pp. 241–263.

[18] Fetterman, D.M., *Ethnography: step by step*, Sage Publications, Newbury Park, Calif., 1989.

[19] Fleiss, J.L., and J. Cohen, "The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability", *Educational and psychological measurement 33*(3), 1973, pp. 613–619.

[20] Gemser, G., C. de Bont, P. Hekkert, and K. Friedman, "Quality perceptions of design journals: The design scholars' perspective", *Design Studies 33*(1), 2012, pp. 4–23.

[21] Giessmann, A., J. Ebermann, and K. Stanoevska-Slabeva, "Do end users accept end user development?", *PACIS*, (2011), 67.

[22] Gilsing, R., O. Ege Adali, O. Turetken, and P. Grefen, "A reference model for the design of service-dominant business models in the smart mobility domain", (2018).

[23] Goffin, K., P. Åhlström, M. Bianchi, and A. Richtnér, "State-of-the-art: The quality of case study research in innovation management", *Journal of Product Innovation Management*, 2019, pp. jpim.12492.

[24] Goldschmidt, G., *Linkography: unfolding the design process*, The MIT Press, Cambridge, MA, 2014.

[25] Gregor, S., and D. Jones, "The anatomy of a design theory", *Journal of the Association for Information Systems 8*(5), 2007, pp. 312–335.

[26] Hay, L., A.H.B. Duffy, C. McTeague, L.M. Pidgeon, T. Vuletic, and M. Grealy, "A systematic review of protocol studies on conceptual design cognition: Design as search and exploration", *Design Science 3*, 2017.

[27] Hevner, A., "A three cycle view of design science research", *Scandinavian Journal of Information Systems 19*(2), 2007.

[28] Hevner, A., S.T. March, J. Park, and S. Ram, "Design science in information systems research", *MIS Quarterly 28*(1), 2004, pp. 75–105.

[29] Holtkamp, P., W. Soliman, and M. Siponen, "Reconsidering the Role of Research Method Guidelines for Qualitative, Mixed-methods, and Design Science Research",

*Proceedings of the 52nd Hawaii International Conference on System Sciences*, (2019), 6280–6289.

[30] Irani, Z., T. Elliman, and P. Jackson, "Electronic transformation of government in the U.K.: A research agenda", *European Journal of Information Systems 16*(4), 2007, pp. 327–335.

[31] Jick, T.D., "Mixing qualitative and quantitative methods: Triangulation in action", *Administrative Science Quarterly 24*(4), 1979, pp. 602–611.

[32] Kemmis, S., R. McTaggart, and R. Nixon, *The Action Research Planner: Doing Critical Participatory Action Research*, Springer, Singapore, 2013.

[33] Klein, H.K., and M.D. Myers, "A set of principles for conducting and evaluating interpretive field studies in information systems", *MIS Quarterly 23*(1), 1999, pp. 67–94.

[34] Knowlton, H.E., "A framework for evaluating inservice workshops", *Teacher Education and Special Education 3*(3), 1980, pp. 58–69.

[35] Landis, J.R., and G.G. Koch, "The measurement of observer agreement for categorical data", *Biometrics 33*(1), 1977, pp. 159–174.

[36] Lewin, K., "Action research and minority problems", *Journal of Social Issues 2*, 1946, pp. 34–46.

[37] Lewins, A., and C. Silver, *Using Software in Qualitative Research: A Step-by-Step Guide*, SAGE Publications, Los Angeles, 2007.

[38] Lomax, H., and N. Casey, "Recording Social Life: Reflexivity and Video Methodology", *Sociological Research Online 3*(2), 1998, pp. 1–26.

[39] March, S.T., and G.F. Smith, "Design and natural science research on information technology", *Decision Support Systems 15*(4), 1995, pp. 251–266.

[40] Morana, S., S. Schacht, A. Scherp, and A. Maedche, "Designing a process guidance system to support user's business process compliance", *ICIS 2014*, (2014).

[41] Nagle, T., and D. Sammon, "The data value map: A framework for developing shared understanding on data initiatives", (2017), 1439–1452.

[42] Oosthuizen, M., "Action research", *Research methods for students, academics and professionals: information management and systems, 2nd edn, Centre for Information Studies, Charles Sturt University, Wagga Wagga*, 2002, pp. 159–75.

[43] Ørngreen, R., and K. Levinsen, "Workshops as a Research Methodology", *Electronic Journal of e-Learning 15*(1), 2017, pp. 70–81.

[44] Peffers, K., T. Tuunanen, M.A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research", *Journal of Management Information Systems 24*(3), 2007, pp. 45–77.

[45] Plowright, D., *Using Mixed Methods*, SAGE Publications Ltd, Los Angeles, 2011.

[46] Ramduny-Ellis, D., A. Dix, P. Rayson, V. Onditi, I. Sommerville, and J. Ransom, "Artefacts as designed, artefacts as used: Resources for uncovering activity dynamics", *Cognition, Technology and Work 7*(2), 2005, pp. 76–87.

[47] Sankar, C., V. Mbarika, and P. Raju, "An Evaluation of a Workshop with a Focus on Fostering Teaching Excellence through Research", *AMCIS 2007 Proceedings*, 2007, pp. 320.

[48] Sein, M.K., O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren, "Action design research", *MIS Quarterly 35*(1), 2011, pp. 37–56.

[49] Siau, K., and M. Rossi, "Evaluation techniques for systems analysis and design modelling methods - a review and comparative analysis", *Information Systems Journal 21*(3), 2011, pp. 249–268.

[50] Sohn, M., and T.-J. Nam, "Understanding the attributes of product intervention for the promotion of pro-environmental behavior: A framework and its effect on immediate user reactions", *International Journal of Design 9*(2), 2015, pp. 55–77.

[51] Storvang, P., B. Mortensen, and A.H. Clarke, "Using workshops in business research: A framework to diagnose, plan, facilitate and analyze workshops", In *Collaborative Research Design: Working with Business for Meaningful Findings*. 2017, 155–174.

[52] Swann, C., "Action research and the practice of design", *Design issues 18*(1), 2002, pp. 49–61.

[53] Thoring, K., R.M. Mueller, and P. Badke-Schaub, "Technology-Supported Design Research", *Proceedings of the 20th International Conference on Engineering Design (ICED 15)*, Design Society (2015), 31–40.

[54] Tremblay, M.C., A.R. Hevner, and D.J. Berndt, "Focus Groups for Artifact Refinement and Evaluation in Design Research", *Communications of the Association for Information Systems 26*(27), 2010, pp. 599–618.

[55] Venable, J., J. Pries-Heje, and R. Baskerville, "FEDS: a Framework for Evaluation in Design Science Research", *European Journal of Information Systems 25*(1), 2016, pp. 77–89.

[56] Venkatesh, V., S.A. Brown, and H. Bala, "Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems", *MIS Quarterly 37*(1), 2013, pp. 21–54.

[57] Villalba, C., A. Jaiprakash, J. Donovan, J. Roberts, and R. Crawford, "Testing literature-based health experience insight cards in a healthcare service co-design workshop", *CoDesign*, 2019.

[58] Whyte, W.F.E., *Participatory action research*, Sage Publications, Inc, Newbury Park, CA, 1991.